Beginning of 01/10/2024

Theorem: D4.1.9

Assume $\mathbb{E}|X| < \infty$ and $\mathbb{E}|Y| < \infty$. Then we have the following properties of conditional:

(1) (Linearity) $\mathbb{E}[aX + Y|\mathcal{F}] = a\mathbb{E}[X|\mathcal{F}] + \mathbb{E}[Y|\mathcal{F}]$.

(2) (Monotonicity) If $X \leqslant Y$ then $\mathbb{E}[X|\mathcal{F}] \leqslant \mathbb{E}[Y|\mathcal{F}]$.

(3) (MCT) If $X_n \geqslant 0$ and $X_n \uparrow X$ (with $\mathbb{E}X < \infty$ then $\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}]$.

Proof. (1) Let $Z = a\mathbb{E}[X|\mathcal{F}] + \mathbb{E}[Y|\mathcal{F}]$. Clearly Z is $\mathcal{F}$-measurable. Then, for any fixed $A \in \mathcal{F}$,

$$\begin{aligned}\int_A Z \, \mathrm{d}\mathbb{P} &= \int_A (a\mathbb{E}[X|\mathcal{F}] + \mathbb{E}[Y|\mathcal{F}] \, \mathrm{d}\mathbb{P} \\ &= a \int_A \mathbb{E}[X|\mathcal{F}] \, \mathrm{d}\mathbb{P} + \int_A \mathbb{E}[Y|\mathcal{F}] \, \mathrm{d}\mathbb{P} \\ &= a \int_A X \, \mathrm{d}\mathbb{P} + \int_A Y \, \mathrm{d}\mathbb{P} = \int_A (aX + Y) \, \mathrm{d}\mathbb{P}.\end{aligned}$$

(2) Let $A \in \mathcal{F}$. Then,

$$\int_A \mathbb{E}[X|\mathcal{F}] \, \mathrm{d}\mathbb{P} = \int_A X \, \mathrm{d}\mathbb{P} \leqslant \int_A Y \, \mathrm{d}\mathbb{P} = \int_A \mathbb{E}[Y|\mathcal{F}] \, \mathrm{d}\mathbb{P}$$

so

$$\int_A (\mathbb{E}[Y|\mathcal{F}] - \mathbb{E}[X|\mathcal{F}]) \, \mathrm{d}\mathbb{P} \geqslant 0.$$

Let $A_n = \{\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[Y|\mathcal{F}] \geqslant 1/n\}$. Then $\mathbb{P}(A_n) = 0$ by applying A_n to the inequality above. Taking unions result in $\mathbb{P}(\mathbb{E}[X|\mathcal{F}] > \mathbb{E}[Y|\mathcal{F}]) = 0$.

(3) Let $Y_n = X - X_n \geqslant 0$. Our assumptions are that $Y_n \downarrow 0$ almost surely, and from (ii) monotonicity,

$$Z_n = \mathbb{E}[Y_n|\mathcal{F}]$$

are decreasing to some limit Z_∞. By linearity, showing $\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}]$ is equivalent to $\mathbb{E}[X - X_n|\mathcal{F}] \downarrow 0$, or $Z_\infty \downarrow 0$. For all $A \in \mathcal{F}$ and n we have

$$\int_A Z_n \, \mathrm{d}\mathbb{P} = \int_A Y_n \, \mathrm{d}\mathbb{P}$$

The LHS in is in particular dominated by $\int_A Z_1 \, \mathrm{d}\mathbb{P}$, and the RHS by $\int_A Y_1 \, \mathrm{d}\mathbb{P}$. Therefore by applying DCT on both sides,

$$\int_A Z_\infty \, \mathrm{d}\mathbb{P} = \int_A 0 \, \mathrm{d}\mathbb{P}.$$

Since Z_∞ is measurable (a limit of measurable functions), $Z_\infty = 0$ a.s. □

Theorem: D4.1.10, Jensen's Inequality

If φ is convex, $\mathbb{E}|X| < \infty$, and $\mathbb{E}|\varphi(X)| < \infty$, then

$$\varphi(\mathbb{E}[X|\mathcal{F}]) \leqslant \mathbb{E}[\varphi(X)|\mathcal{F}].$$

(We require $\varphi(X) \in L^1$ so the RHS is well-defined.)

Proof. We assume that φ is not a linear function (for otherwise the result is apparent). Define

$$S := \{(a,b) : a,b \in \mathbb{Q}, ax + b \leqslant \varphi(x) \text{ for all } x \in \mathbb{R}\}$$

the collection of all rationally parametrized lines in the hypograph of φ. We observe that if φ is not linear, then $\varphi(x) = \sup\{ax + b : (a,b) \in S\}$ (pointwise). It follows (from this un-proven observation) that

$$\mathbb{E}[\varphi(X)|\mathcal{F}] \geqslant a\mathbb{E}[X|\mathcal{F}] + b \text{ a.s.,} \qquad \text{for all } (a,b) \in S.$$

Taking supremum,

$$\sup_{(a,b)\in S} a\mathbb{E}[X|\mathcal{F}] + b \leqslant \mathbb{E}[\varphi(X)|\mathcal{F}] \text{ a.s.}$$

We need to define S only on rationals because the above inequality only necessarily holds over a countable set. The LHS is $\varphi(\mathbb{E}[X|\mathcal{F}])$ so we are done. □

Corollary: D4.1.11

Fix $p \geqslant 1$ and let $X \in L^p$. Then $|\mathbb{E}[X|\mathcal{F}]|^p \leqslant \mathbb{E}[|X|^p|\mathcal{F}]$. *Proof: use Jensen's with* $\varphi(x) = |x|^p$.

A result from this inequality:

$$\mathbb{E}\Big[|\mathbb{E}[X|\mathcal{F}]|^p\Big] \leqslant \mathbb{E}\Big[\mathbb{E}[|X|^p|\mathcal{F}]\Big] = \mathbb{E}[|X|^p]$$

(where the last equality follows from letting $A = \Omega$ in the defining property (ii) of conditionals), so performing a conditional expectation in L^p space is a contraction.

Theorem: D4.1.12

If $\mathcal{F} \subset \mathcal{G}$ and $\mathbb{E}[X|\mathcal{G}] \in \mathcal{F}$, then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X|\mathcal{G}]$.

Proof. Condition (i) already holds since $\mathbb{E}[X|\mathcal{G}] \in \mathcal{F}$. To check condition (ii): we know

$$\int_A \mathbb{E}[X|\mathcal{G}] \, \mathrm{d}\mathbb{P} = \int_A X \, \mathrm{d}\mathbb{P}$$

for every $A \in \mathcal{G}$ by the defining properties of $\mathbb{E}[X|\mathcal{G}]$. And of course it holds for $A \in \mathcal{F}$. □

Theorem: D4.1.13, Tower Property of Conditional Expectations

If $\mathcal{F}_1 \subset \mathcal{F}_2$, then

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{F}_1]|\mathcal{F}_2\right] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1\right] = \mathbb{E}[X|\mathcal{F}_1].$$

"Smaller σ-algebra always wins." In the first term, in the inner bracket we are averaging over less information, so we already have lost some information by the time we get to the outer bracket. In the second quantity we indeed had more information in our first averaging process, but in the outer bracket we discarded some of those.

Proof. We define $Y = \mathbb{E}[X|\mathcal{F}_1]$. We already know $\mathbb{E}[Y|\mathcal{F}_2] = Y$ since $\mathcal{F}_1 \subset \mathcal{F}_2$, which implies $Y \in \mathcal{F}_2$. All the way in the beginning (D4.1.3) we showed that in the case of *perfect information*, the "best guess" is just the variable itself: $\mathbb{E}[Y|\mathcal{F}_2] = Y$.

For the other equality, we know that $\mathbb{E}[X|\mathcal{F}_1] \in \mathcal{F}_1$ and that for each $A \in \mathcal{F}_1 \subset \mathcal{F}_2$,

$$\int_A \mathbb{E}[X|\mathcal{F}_2]\, \mathrm{d}\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{F}_1]\, \mathrm{d}\mathbb{P} = \int_A X\, \mathrm{d}\mathbb{P}.$$

□

Theorem: D4.1.14

If $X \in \mathcal{F}$, $\mathbb{E}|Y| < \infty$, and $\mathbb{E}|XY| < \infty$, then

$$\mathbb{E}[XY|\mathcal{F}] = X\mathbb{E}[Y|\mathcal{F}].$$

One last theorem for conditional expectation. We now quantify the "best guess:"

Theorem: D4.1.15

Suppose $\mathbb{E}X^2 < \infty$. Then $\mathbb{E}[X|\mathcal{F}]$ is the variable $Y \in \mathcal{F}$ that uniquely minimizes $\mathbb{E}[(X - Y)^2]$.