Lecture 6 · 30.4.2021 · 10:15–12:00 via Zoom

Conditional expectation (continuation)

Definition. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $\mathscr{G} \subseteq \mathscr{F}$ a sub-σ-algebra of $\mathscr{F}$ and let $X: \Omega \to \mathbb{R}$ be a (real) integrable random variable (i.e. $\mathbb{E}[|X|] < \infty$). The a conditional expectation of X with respect to the σ-algebra $\mathscr{G}$ (or given $\mathscr{G}$) is a real integrable random variable $Z: \Omega \to \mathbb{R}$ such that*

1. *Z is $\mathscr{G}$-measurable;*
2. *The equation*
$$\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Z \mathbb{1}_A] \tag{1}$$
holds for all $A \in \mathscr{G}$.

Remark. We can rewrite (1) as

$$\int_A X(\omega) \mathbb{P}(\mathrm{d}\omega) = \int_A Z(\omega) \mathbb{P}(\mathrm{d}\omega)$$

for all $A \in \mathscr{G}$. This means that we cannot see any difference between X and Z if the only things we know are the events in $\mathscr{G}$.

Today we are going to prove existence and give some of the basic properties of cond. exp. Many of these properties will be the subject of the next exercise sheet. Many of the proof, especially at the beginning must use directly the definition.

An important special case of cond. exp. is when $\mathscr{G} = \sigma(Y)$, i.e. the conditioning σ-algebra is generated by a random variable $Y: \Omega \to (E, \mathscr{E})$ with values in an arbitrary measure space $(E, \mathscr{E})$. In this case the cond. expectation Z given $\mathscr{G} = \sigma(Y)$ has a specific form.

Theorem 1. *Let $X: (\Omega, \mathscr{F}) \to (\Theta, \mathscr{H})$ be a r.v. with values in $(\Theta, \mathscr{H})$ and measurable wrt. $\mathscr{F}$ (as usual). Let $Y: (\Omega, \sigma(X)) \to (\Upsilon, \mathscr{G})$ a r.v. which is $\sigma(X)$-measurable and with values in the measure space $(\Upsilon, \mathscr{G})$.*
Then there exists a measurable map $h: (\Theta, \mathscr{H}) \to (\Upsilon, \mathscr{G})$ such that $Y = h \circ X = h(X)$. That is the following diagram commutes:

$$\begin{array}{ccc} (\Omega, \sigma(X)) & \xrightarrow{X} & (\Theta, \mathscr{H}) \\ {}_{Y}\searrow & & \swarrow_{h} \\ & (\Upsilon, \mathscr{G}) & \end{array}$$

The theorem says that any r.v. Y which is $\sigma(X)$ measurable is actually just a (measurable) function of X, i.e. $Y = h(X)$.

Proposition 2. *If $Z: \Omega \to \mathbb{R}$ is a cond. expectation of an integrable $X: \Omega \to \mathbb{R}$ given $\sigma(Y)$ for some r.v. $Y: \Omega \to (E, \mathscr{E})$ then there exists a measurable function $h: (E, \mathscr{E}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that*

$$Z = \mathbb{E}[X|\sigma(Y)] = h(Y).$$

Remark that this is indeed what happens in the elementary situation in the last lecture.

Since this situation is quite common we have a specific notation, given two random variables $X\colon \Omega \to \mathbb{R}$ and $Y\colon \Omega \to (E, \mathscr{E})$ then we define the cond. exp. of X given Y as

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)].$$

In particular if $Y\colon \Omega \to (E, \mathscr{E}), Y'\colon \Omega \to (E', \mathscr{E}')$ are two r.v. such that $\sigma(Y) = \sigma(Y')$ then

$$\mathbb{E}[X|Y] = \mathbb{E}[X|Y']$$

almost surely.

This happens for example when $Y = f(Y')$ with $f\colon (E', \mathscr{E}') \to (E, \mathscr{E})$ (measurable) bijective.

Exercise: Can we in this case prove that indeed $\sigma(Y) = \sigma(Y')$ or we need more conditions?

> **Remark.** In general we can define cond. exp. only for integrable r.v. (because there is an integral involved in the definition). However if $X \geqslant 0$ then we can always define the conditional expectation since the integrals in the eq. (1) are always well defined since it is not restrictive to define in this case the cond. exp. as a positive r.v..

Existence

The goal is to prove the following theorem (introduce in the last lecture)

Theorem 3. *For any $X\colon \Omega \to \mathbb{R}$ integrable and $\mathscr{G} \subseteq \mathscr{F}$ the cond. exp. $\mathbb{E}[X|\mathscr{G}]$ exists.*

There are two main proofs of this theorem. We will use one which goes via an intermediate result involving the condition that X is square-integrable, this means that $X \in \mathscr{L}^2(\Omega, \mathscr{F}, \mathbb{P})$, i.e. $\mathbb{E}[|X|^2] < \infty$. In this case we can use the geometry of the Hilbert space L^2 to perform the construciton of the cond. exp. After that we remove the additional integrability condition via an approximation argument.

Recall that $L^2(\Omega, \mathscr{F}, \mathbb{P})$ is a real Hilbert space with scalar product

$$\langle X, Y \rangle = \mathbb{E}[XY], \qquad X, Y \in L^2(\Omega, \mathscr{F}, \mathbb{P}).$$

Let us observe preliminarly the following: *assume* $X \in L^2$ and Y is a cond. exp. for X given $\mathscr{G}$ and that $Y \in L^2$ (this is not automatic for what we now).

Then for any $Z\colon \Omega \to \mathbb{R}$ bounded and measurable wrt $\mathscr{G}$ then one can check that

$$\begin{aligned}\mathbb{E}[|X - Z|^2] &= \mathbb{E}[X^2] - 2\mathbb{E}[XZ] + \mathbb{E}[Z^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[YZ] + \mathbb{E}[Z^2] \\ &= \mathbb{E}[|Y - Z|^2] + \mathbb{E}[|X - Y|^2] + \underbrace{2\mathbb{E}[XY] - 2\mathbb{E}[Y^2]}_{=0} = \mathbb{E}[|Y - Z|^2] + \mathbb{E}[|X - Y|^2]\end{aligned}$$

where we used that $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ and we also used that

$$\mathbb{E}[(X - Y)Y] = 0,$$

indeed we have

$$\mathbb{E}[X\phi_M(Y)] = \mathbb{E}[Y\phi_M(Y)] \tag{2}$$

with $\phi_M(x) = (-M \wedge (x \vee M))$ is a function which is bounded, so $\phi_M(Y)$ is bounded for any $M > 0$ and therefore the equality holds by definition of cond. exp since now $\phi_M(Y)$ is a generic bounded r.v. which is $\mathscr{G}$ measurable since Y is. Then in eq. (2) one can take the limit as $M \to \infty$ using the fact that both X, Y are in L^2. For example observe that

$$|\mathbb{E}[X\phi_M(Y)] - \mathbb{E}[XY]| = |\mathbb{E}[X(\phi_M(Y) - Y)]| \leqslant \mathbb{E}[|X(\phi_M(Y) - Y)|]$$

$$\leqslant (\mathbb{E}[|X|^2])^{1/2}(\mathbb{E}[|\phi_M(Y) - Y|^2])^{1/2} \to 0$$

by dominated convergence since $|\phi_M(Y) - Y|^2 \leqslant 2|Y|^2$ which is integrable. So one can show in this way that for $X, Y \in L^2$ we have indeed

$$\mathbb{E}[XY] = \mathbb{E}[YY].$$

From the point of view of Hilbert space this means that $\langle X - Y, Y\rangle = 0$ i.e. $X - Y$ is orthogonal to Y. In particular we have proven that

$$\mathbb{E}[|X - Z|^2] = \mathbb{E}[|Y - Z|^2] + \mathbb{E}[|X - Y|^2]$$

so also that $Y - Z$ and $X - Y$ are orthogonal for any Z which is $\mathscr{G}$ measurable and bounded. By the same approximations argument as above we have also this inequality for all $Z \in L^2(\Omega, \mathscr{G}, \mathbb{P}) = L^2(\mathscr{G})$, i.e. $\mathscr{G}$ measurable and square-integrable random variables.

So

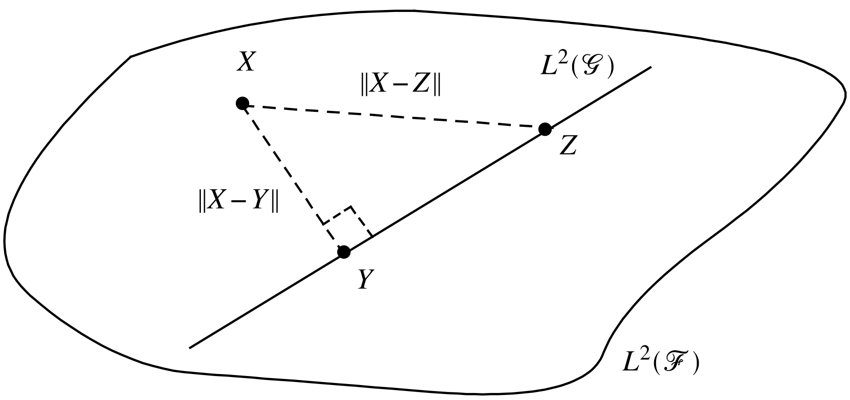
$$\mathbb{E}[|X - Y|^2] = \inf_{Z \in L^2(\mathscr{G})} \mathbb{E}[|X - Z|^2] \tag{3}$$

and the infimum is obtained for $Z = Y$ assuming the cond. exp. Y is indeed in $L^2(\mathscr{G})$.

Note that $L^2(\mathscr{G}) \subseteq L^2(\mathscr{F})$ and moreover that it is closed subspace with the topology induced by the L^2 norm (by completeness of $L^2(\mathscr{G})$ which is just a sub Hilbert space of $L^2(\mathscr{F})$).

Eq. (3) tells us that if the cond. exp. exists and it is in $L^2(\mathscr{G})$, **then** it must be a minimizer of the expression $\mathbb{E}[|X - Z|^2]$ with Z running over all the elements of $L^2(\mathscr{G})$.

> The conditional expectation is the *best approximation* of $X \in L^2(\mathscr{F})$ by an element of $L^2(\mathscr{G})$ according to the L^2 distance.



This hints to a concrete strategy to construct the cond. exp., namely looking for infima of this functional.

Lemma 4. *Fix $\mathscr{G} \subseteq \mathscr{F}$, then $L^2(\mathscr{G})$ is a closed vector subspace of $L^2(\mathscr{F})$ and for all $X \in L^2(\mathscr{F})$ there exists a unique $Y \in L^2(\mathscr{G})$ such that*

1. $\mathbb{E}[|X-Y|^2] = \inf_{Z \in L^2(\mathscr{G})} \mathbb{E}[|X-Z|^2]$
2. $X - Y \perp L^2(\mathscr{G})$.

We call this Y the orthogonal projection of X onto $L^2(\mathscr{G})$.

Proof. The vector space $L^2(\mathscr{G})$ is complete wrt. the L^2 norm and therefore is also closed in $L^2(\mathscr{F})$ because the norm (and therefore the metric topology) is the same. We have now to prove the existence of Y with the two properties. Let

$$\Delta = \inf_{Z \in L^2(\mathscr{G})} \mathbb{E}[|X-Z|^2] \geqslant 0.$$

I can consider a sequence $(Y_n)_{n \geqslant 0} \subseteq L^2(\mathscr{G})$ of almost minimizers of this quantity, that is

$$\mathbb{E}[|X-Y_n|^2] \leqslant \Delta + \frac{1}{n}$$

as $n \to \infty$. I want to prove that this sequence is Cauchy. We have

$$\mathbb{E}[|X-Y_n|^2] + \mathbb{E}[|X-Y_m|^2] = 2\,\mathbb{E}[|X-(Y_n+Y_m)/2|^2] + \frac{1}{2}\mathbb{E}[|Y_n-Y_m|^2]$$

(using $\mathbb{E}[|A+B|^2] + \mathbb{E}[|A-B|^2] = 2\mathbb{E}[A^2] + 2\mathbb{E}[B^2]$). Now observe that $(Y_n+Y_m)/2 \in L^2(\mathscr{G})$, which gives that

$$\frac{1}{2}\mathbb{E}[|Y_n-Y_m|^2] \leqslant \mathbb{E}[|X-Y_n|^2] + \mathbb{E}[|X-Y_m|^2] - 2\Delta \leqslant \frac{1}{n} + \frac{1}{m} \to 0$$

as $n, m \to \infty$ which shows that $(Y_n)_{n \geqslant 0} \subseteq L^2(\mathscr{G})$ is Cauchy. Therefore there exists $Y \in L^2(\mathscr{G})$ such that $Y_n \to Y$ in $L^2(\mathscr{G})$ and therefore in $L^2(\mathscr{F})$ and is easy to check that

$$\mathbb{E}[|X-Y|^2] = \Delta.$$

Take $t \in \mathbb{R}$ and $H \in L^2(\mathscr{G})$ and consider $Z = Y + tH \in L^2(\mathscr{G})$, then

$$\Delta \leqslant \mathbb{E}[|X-Y-tH|^2] = \underbrace{\mathbb{E}[|X-Y|^2]}_{=\Delta} + 2t\mathbb{E}[(X-Y)H] + t^2\mathbb{E}[H^2]$$

since this must hold for any $t \in \mathbb{R}$ by taking t small this implies that $\mathbb{E}[(X-Y)H] = 0$ (it is the derivative in zero of a positive polynomial). This means that

$$\langle X-Y, H \rangle = \mathbb{E}[(X-Y)H] = 0$$

for all $H \in L^2(\mathscr{G})$ which is the required orthogonality condition. Uniqueness of Y follows from the same argument we used for conditional expectation (exercise). □

This settles the existence problem when $X \in L^2(\mathscr{F})$ and in this case we can take $\mathbb{E}[X|\mathscr{G}] \in L^2(\mathscr{G})$ to be the orthogonal projection of X in $L^2(\mathscr{G})$. More generally we can proceed by approximation.

Theorem 5. *For all $X \in L^1(\mathscr{F})$ and σ-algebra $\mathscr{G} \subseteq \mathscr{F}$ the conditional expectation $\mathbb{E}[X|\mathscr{G}]$ exists.*

Proof. Let's assume first that $X \geqslant 0$. Let $X_n = (X \wedge n) \in L^2(\mathscr{F})$ for any $n \geqslant 0$. In this case we proved that $Y_n = \mathbb{E}[X_n|\mathscr{G}]$ exists. The first thing to realize is that $Y_m \geqslant Y_n$ for any $m \geqslant n$. Indeed by definition we have for all $A \in \mathscr{G}$

$$\mathbb{E}[\mathbb{1}_A(Y_m - Y_n)] = \mathbb{E}\big[\mathbb{1}_A \underbrace{(X_m - X_n)}_{\geqslant 0}\big] \geqslant 0$$

and choosing $A = A_k = \{Y_m - Y_n < k^{-1}\} \in \mathscr{G}$ one deduce that $\mathbb{P}(A_k) = 0$ for any $k = 1, \ldots$ which implies that $\mathbb{P}(\cup_k A_k) = 0$ that is $Y_m \geqslant Y_n$ a.s.. So $(Y_n)_{n \geqslant 1}$ is almost surely increasing and therefore has a pointwise limit $Y = \lim_n Y_n = \sup_n Y_n$ (see in the notes the details).

One now checks that $\mathbb{E}[\mathbb{1}_A Y] = \mathbb{E}[\mathbb{1}_A X]$ by taking monotone limits in $\mathbb{E}[\mathbb{1}_A Y_n] = \mathbb{E}[\mathbb{1}_A X_n]$. This gives $\mathbb{E}[X|\mathscr{G}] = Y$ when $X \geqslant 0$.

Then one extends to arbitrary X by decomposing into positive and negative parts $X = X_+ - X_-$, by letting

$$\mathbb{E}[X|\mathscr{G}] := \mathbb{E}[X_+|\mathscr{G}] - \mathbb{E}[X_-|\mathscr{G}].$$

□
