Lecture 7 · 4.5.2021 · 14:15–16:00 via Zoom

Conditional expectation (end)

Last week we proved existence of cond. exp for L^2 random variables via orthogonal projection in the Hilbert space L^2 and then extended it to all r.v. in L^1 via usual arguments of measure theory, namely monotone approximation for positive r.v. and then decomposition into positive and negative parts for general integrable r.v.

Warning: $\mathbb{E}[X], \mathbb{E}[X|\mathcal{G}]$ are two very different objects. $\mathbb{E}[X]$ is a number giving the result of computing an integral. $\mathbb{E}[X|\mathcal{G}]$ it is a random variabl with certain properties.

Properties of conditional expectation

Proposition. *For all $X, Y \in L^1(\mathcal{F})$ and all $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ (sub-σ-algebras of $\mathcal{F}$) we have the following properties of conditional expectation (all valid **only** $\mathbb{P}$-a.s.):*

a) Linearity: for all $\lambda, \mu \in \mathbb{R}$ we have

$$\mathbb{E}[\lambda X + \mu Y|\mathcal{G}](\omega) = \lambda \mathbb{E}[X|\mathcal{G}](\omega) + \mu \mathbb{E}[Y|\mathcal{G}](\omega);$$

(in particular $\mathbb{E}[\lambda|\mathcal{G}] = \lambda$)

b) Positivity: for any $X \geqslant 0$ $\mathbb{P}$-a.s. we have

$$\mathbb{E}[X|\mathcal{G}] \geqslant 0;$$

c) Monotone convergence: for any non-decreasing sequence $(X_n)_{n\geqslant 1}$ of integrable r.v. such that $X = \lim_n X_n = \sup_n X_n$ we have

$$\mathbb{E}[X|\mathcal{G}] = \sup_n \mathbb{E}[X_n|\mathcal{G}].$$

d) Jensen's inequality. For any $\varphi: \mathbb{R} \to \mathbb{R}$ convex and such that $\varphi(X) \in L^1$ we have

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geqslant \varphi(\mathbb{E}[X|\mathcal{G}])$$

(to prove this use that $\varphi(x)$ can be bounded below by a suitable straight line, see the proof for the standard expectation)

e) Contractivity in L^p with $p \geqslant 1$: if $X \in L^p$ then $\mathbb{E}[X|\mathcal{G}] \in L^p$ and

$$\|\mathbb{E}[X|\mathcal{G}]\|_{L^p} \leqslant \|X\|_{L^p}.$$

f) Telescoping: If $\mathcal{H} \subseteq \mathcal{G}$ then the smallest σ-algebra wins:

$$\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$$

(in general one has $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] \neq \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$).

g) *If $Z \,\hat{\in}\, \mathscr{G}$ (i.e. measurable wrt. $\mathscr{G}$), $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|XZ|] < \infty$ then*

$$\mathbb{E}[XZ|\mathscr{G}] = Z\,\mathbb{E}[X|\mathscr{G}],$$

you can take out of the cond. exp. any $\mathscr{G}$ measurable r.v. in particular $\mathbb{E}[Z|\mathscr{G}] = Z$.

Remark. We will use the notation $Z \,\hat{\in}\, \mathscr{G}$ to denote that the r.v. Z is $\mathscr{G}$ measurable.

An important lemma on the relation between cond. exp. and UI.

Lemma 1. *Let $X \in L^1(\mathscr{F})$ and for any $\mathscr{G} \subseteq \mathscr{F}$ define $X_{\mathscr{G}} = \mathbb{E}[X|\mathscr{G}]$. Then the family*

$$\mathscr{X} = \{X_{\mathscr{G}} : \mathscr{G} \subseteq \mathscr{F}\}$$

is an uniformly integrable family of random variables.

Proof. Recall UI: we have to prove that for any $\varepsilon > 0$ there exists $L > 0$ such that

$$\sup_{X_{\mathscr{G}} \in \mathscr{X}} \mathbb{E}[|X_{\mathscr{G}}| \mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}] \leqslant \varepsilon.$$

Observe that $X_{\mathscr{G}} \,\hat{\in}\, \mathscr{G}$ and therefore $\{|X_{\mathscr{G}}| \geqslant L\} \in \mathscr{G}$, as a consequence

$$\mathbb{E}[|X_{\mathscr{G}}| \mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}] = \mathbb{E}[|\mathbb{E}[X|\mathscr{G}]| \mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}] \underset{\text{Jensen}}{\leqslant} \mathbb{E}\big[\mathbb{E}[|X||\mathscr{G}] \underbrace{\mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}}_{\hat{\in}\mathscr{G}}\big]$$

$$\leqslant \mathbb{E}[\mathbb{E}[|X|\,\mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}|\mathscr{G}]] = \mathbb{E}[|X|\,\mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}]$$

since $\mathbb{E}[\mathbb{E}[Y|\mathscr{G}]] = \mathbb{E}[\mathbb{E}[Y|\mathscr{G}]\mathbb{1}_{\Omega}] = \mathbb{E}[Y\mathbb{1}_{\Omega}] = \mathbb{E}[Y]$ using the definition since $\Omega \in \mathscr{G}$. We have now

$$\mathbb{E}[|X_{\mathscr{G}}| \mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}] \leqslant \mathbb{E}[|X|\,\mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}].$$

Since the r.v. X is UI (since any integrable r.v. is) we have that there exists $\delta > 0$ so that for any $A \in \mathscr{F}$ with $\mathbb{P}(A) \leqslant \delta(\varepsilon)$ we have $\mathbb{E}[|X|\,\mathbb{1}_A] \leqslant \varepsilon$.
Then it suffices to take $L = L(\delta)$ large so that

$$\mathbb{P}(|X_{\mathscr{G}}| \geqslant L) \underset{\text{Markov}}{\leqslant} \frac{\mathbb{E}[|X_{\mathscr{G}}|]}{L} \underset{\text{Contractivity in } L^1}{\leqslant} \frac{\mathbb{E}[|X|]}{L} \leqslant \delta$$

to finally have that for $L = L(\delta(\varepsilon))$ we have

$$\mathbb{E}[|X_{\mathscr{G}}| \mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}] \leqslant \mathbb{E}[|X|\,\mathbb{1}_{|X_{\mathscr{G}}| \geqslant L}] \leqslant \varepsilon$$

independently of $\mathscr{G}$. This proves UI of the family $\mathscr{X}$. □

Relations with independence

Recall the notion of independence: two events A, B are independent wrt. $\mathbb{P}$ if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Generalisations involving families of σ-algebras or rand. vars. are also possible.

Definition.

a) *A family* $(\mathscr{A}_i)_{i\in I}$ *of sub-*σ*-algebras of* $\mathscr{F}$ *are independent iff for any choice of* $J\subseteq I$ *finite and any* $A_i\in\mathscr{A}_i$*,* $i\in J$ *we have*

$$\mathbb{P}(\cap_{j\in J}A_j)=\prod_{j\in J}\mathbb{P}(A_j).$$

(pair-wise independence is not sufficient for general independence)

b) *We say that a r.v.* X *is independent from a* σ*-algebra* $\mathscr{G}$ *if* $\{\sigma(X),\mathscr{G}\}$ *are independent.*

c) *A family of r.v.* $(X_i)_{i\in I}$ *is independent if the family* $(\sigma(X_i))_{i\in I}$ *of* σ*-algebras is independent.*

Proposition.

a) *If* $X\in L^1(\mathscr{F})$ *is independent of* $\mathscr{G}$ *then*

$$\mathbb{E}[X|\mathscr{G}]=\mathbb{E}[X].$$

b) *If* $\mathscr{H},\mathscr{G}$ *are independent and* $\mathscr{G}'\subseteq\mathscr{G}$ *and* $X\in L^1(\mathscr{G})$ *then*

$$\mathbb{E}[X|\mathscr{H},\mathscr{G}']=\mathbb{E}[X|\mathscr{G}']$$

(where $\mathbb{E}[X|\mathscr{H},\mathscr{G}']:=\mathbb{E}[X|\sigma(\mathscr{H},\mathscr{G}')]$*). That is we can ignore additional independent information in the conditioning.*

c) *If* $X_1,\ldots,X_n$ *is a finite family of real independent r.vs. and* $f(X_1,\ldots,X_n)\in L^1(\mathscr{F})$ *then*

$$\mathbb{E}[f(X_1,\ldots,X_n)|X_1]=\varphi(X_1)$$

where the function φ *is explicitly given by*

$$\varphi(x)=\mathbb{E}[f(x,X_2,\ldots,X_n)],\qquad x\in\mathbb{R}.$$

(Note that $\varphi(X_1)\neq\mathbb{E}[f(X_1,X_2,\ldots,X_n)]$*) With another more detailed notation we have*

$$\mathbb{E}[f(X_1,\ldots,X_n)|X_1](\omega)=\varphi(X_1(\omega))=\int_\Omega f(X_1(\omega),X_2(\omega'),\ldots,X_n(\omega'))\mathbb{P}(\mathrm{d}\omega').$$

Proof. a) Exercise.

b) We can assume that $X\geqslant 0$ (the general case can be handled by decomposition). Let $G\in\mathscr{G}'$ and $H\in\mathscr{H}$, by definition of cond. exp:

$$\mathbb{E}[\mathbb{E}[X|\mathscr{H},\mathscr{G}']\mathbb{1}_G\mathbb{1}_H]=\mathbb{E}[X\mathbb{1}_G\mathbb{1}_H]$$

By independence of $\mathscr{G}$ and $\mathscr{H}$:

$$\mathbb{E}[X\mathbb{1}_G\mathbb{1}_H]=\mathbb{E}[X\mathbb{1}_G]\,\mathbb{E}[\mathbb{1}_H]$$

and by definition of $\mathbb{E}[X|\mathscr{G}']$ we have

$$\mathbb{E}[X\mathbb{1}_G]\,\mathbb{E}[\mathbb{1}_H]=\mathbb{E}[\mathbb{E}[X|\mathscr{G}']\mathbb{1}_G]\,\mathbb{E}[\mathbb{1}_H].$$

By using independence of $\mathscr{G}$ and $\mathscr{H}$ again we have

$$\mathbb{E}[\mathbb{E}[X|\mathscr{G}']\mathbb{1}_G]\,\mathbb{E}[\mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}']\mathbb{1}_G\mathbb{1}_H].$$

Therefore

$$\mathbb{E}[X\mathbb{1}_{G\cap H}] = \mathbb{E}[\mathbb{E}[X|\mathscr{H},\mathscr{G}']\mathbb{1}_{G\cap H}] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}']\mathbb{1}_{G\cap H}].$$

To conclude it is enough to show that this equality is valid when we replace $G\cap H$ by any element of $\sigma(\mathscr{G}',\mathscr{H})$. This is the point where we can use the monotone class theorem: the property expressed by the above equality is true for $\mathbb{1}_{G\cap H}$, it is linear and pass to the monotone limits (because $X\geqslant 0$ and therefore also $\mathbb{E}[X|\mathscr{G}']\geqslant 0$). So we conclude it holds for all the sets in the σ-algebra generated by $\mathscr{G}'\cap\mathscr{H}$, namely $\sigma(\mathscr{G}',\mathscr{H})$.

c) To prove the explicit form of φ just use Fubini theorem on the joint law of X_1 and $(X_2,\ldots,X_n)$. Actually consider the case $n=2$ is sufficient for a general proof. Note that checking the definition of cond. exp. in this case is equivalent to check that for any bounded and measurable $h\colon\mathbb{R}\to\mathbb{R}$ we have

$$\mathbb{E}[h(X_1)f(X_1,\ldots,X_n)] = \mathbb{E}[h(X_1)\varphi(X_1)].$$

Indeed recall that any $\sigma(X_1)$-measurable r.v. Z has the form $Z=h(X_1)$. By definition of expectation we have

$$\begin{aligned}
\mathbb{E}[Zf(X_1,\ldots,X_n)] &= \mathbb{E}[h(X_1)f(X_1,\ldots,X_n)]\\
&= \int_{\mathbb{R}^n} h(x_1)f(x_1,x_2,\ldots,x_n)\prod_{i=1}^n \mathbb{P}_{X_i}(\mathrm{d}x_i)\\
&= \int_{\mathbb{R}} h(x_1)\underbrace{\left[\int_{\mathbb{R}^{n-1}} f(x_1,x_2,\ldots,x_n)\prod_{i=2}^n \mathbb{P}_{X_i}(\mathrm{d}x_i)\right]}_{\varphi(x_1)=\mathbb{E}[f(x_1,X_2,\ldots,X_n)]}\mathbb{P}_{X_1}(\mathrm{d}x_1)\\
&= \int_{\mathbb{R}} h(x_1)\varphi(x_1)\mathbb{P}_{X_1}(\mathrm{d}x_1) = \mathbb{E}[h(X_1)\varphi(X_1)]\\
&= \mathbb{E}[Z\varphi(X_1)]
\end{aligned}$$

this holds for any $Z\,\hat{\in}\,\sigma(X_1)$ and therefore we can conclude that $\mathbb{E}[f(X_1,\ldots,X_n)|X_1]=\varphi(X_1)$.

□

Example. Let $(X_i)_{i=1,\ldots,n}$ a vector of i.i.d. integrable random variables and let

$$S=\sum_{i=1}^n X_i.$$

We want to compute $\mathbb{E}[X_1|S]$. The first observation is that there must exist a measurable function $g\colon\mathbb{R}\to\mathbb{R}$ such that

$$\mathbb{E}[X_1|S] = \mathbb{E}[X_k|S] = g(S)$$

for any $k=1,\ldots,n$. The function g is independent of the index of the variable: intuitively no variable can be distinguished from each other. Indeed by definition we must have

$$\mathbb{E}[X_1h(S)] = \mathbb{E}[g(S)h(S)]$$

for any bounded measurable function $h\colon \mathbb{R} \to \mathbb{R}$. However

$$\mathbb{E}[X_1 h(S)] = \mathbb{E}[X_1 h(X_1 + \cdots + X_n)] = \mathbb{E}[X_{\sigma(1)} h(X_{\sigma(1)} + \cdots + X_{\sigma(n)})]$$

where $\sigma \in S_n$ is a permutation of $\{1, \ldots, n\}$. This is true since the law of the vector $(X_1, \ldots, X_n)$ is equal to the law of the vector $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ and coincide with a product measure on n equal measures

$$\mathbb{P}_{(X_{\sigma(1)}, \ldots, X_{\sigma(n)})} = \mathbb{P}_{X_{\sigma(1)}} \otimes \cdots \otimes \mathbb{P}_{X_{\sigma(n)}} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_1} = (\mathbb{P}_{X_1})^{\otimes n}.$$

We say in this case that the vector $(X_1, \ldots, X_n)$ is **exchangeable**, i.e. its law is invariant under permutations. Therefore by choosing σ appropriately we have

$$\mathbb{E}[X_{\sigma(1)} h(X_{\sigma(1)} + \cdots + X_{\sigma(n)})] = \mathbb{E}[X_{\sigma(1)} h(X_1 + \cdots + X_n)]$$

$$= \mathbb{E}[X_k h(X_1 + \cdots + X_n)] = \mathbb{E}[X_k h(S)]$$

which implies that

$$\mathbb{E}[X_1 | S] = \mathbb{E}[X_k | S]$$

for any $k = 1, \ldots, n$.

Now by linearity we have

$$S = \mathbb{E}[S|S] = \mathbb{E}[X_1 + \cdots + X_n | S] = \mathbb{E}[X_1|S] + \cdots + \mathbb{E}[X_n|S] = n\,\mathbb{E}[X_1|S] = n\,g(S)$$

as a consequence we have proven that $g(S) = S/n$ and in particular

$$\mathbb{E}[X_1|S] = \frac{S}{n}.$$