Igor Kortchemski
Probality theory,
Autumn 2023
ETH

Chapter 4: Part 1, Conditional expectation

Outline: 1) The discrete setting
2) Definition and first properties
3) Non-negative r.v.
4) Convergence theorems
5) Some useful properties
6) Conditional density functions

Intuitively speaking, the goal is to see how the knowledge of information (i.e a σ-field) modifies probability measures. Here we shall define the conditional expectation of random variables given a σ-field.

1) The discrete setting

We first condition with respect to an event.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $B \in \mathcal{A}$ with $\mathbb{P}(B) > 0$. We can define $\mathbb{P}(\cdot \mid B)$ the so-called conditional probability given B by $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ for every $A \in \mathcal{A}$

Similarly, for $X \in L^1$, we define $\mathbb{E}[X \mid B] = \dfrac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}$

<u>Interpretation</u>: it is the average value of X when B occurs.

The notation $\mathbb{E}[X \mid B]$ comes from the following fact:

Lemma Let $X : (\Omega, \mathcal{A}) \to \mathbb{R}_+$ be a random variable. Its expectation with respect to $\mathbb{P}(\cdot \mid B)$ is $\dfrac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}$.

If we set $\tilde{\mathbb{P}}(A) = \mathbb{P}(A \mid B)$ for $A \in \mathcal{A}$, the expectation of X with respect to $\mathbb{P}(\cdot \mid B)$ is by definition the integral of X with respect to $\tilde{\mathbb{P}}$, that is the quantity $\int_\Omega X(\omega)\, \tilde{\mathbb{P}}(d\omega)$.

Proof We show that $\int_\Omega X(\omega)\, \tilde{\mathbb{P}}(d\omega) \overset{(*)}{=} \dfrac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}$

Step 1 for $X = \mathbb{1}_A$, $\int_\Omega \mathbb{1}_A(\omega)\, \tilde{\mathbb{P}}(d\omega) = \tilde{\mathbb{P}}(A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ which is $\dfrac{\mathbb{E}[\mathbb{1}_A \mathbb{1}_B]}{\mathbb{P}(B)}$

Step 2 By linearity $(*)$ holds for every simple random variable (a r.v. taking a finite number of values)

Step 3 Let $0 \leq X_n \uparrow X$ with X_n simple. Then

$$\int_\Omega X(\omega)\, \tilde{\mathbb{P}}(d\omega) \xleftarrow[n\to\infty]{} \int_\Omega X_n(\omega)\, \tilde{\mathbb{P}}(d\omega) = \frac{\mathbb{E}[X_n \mathbb{1}_B]}{\mathbb{P}(B)} \xrightarrow[n\to\infty]{} \frac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}$$

by monotone convergence (twice)

(N.B. Proof not done during the lecture)

We shall now condition with respect to a discrete random variable

Now let $Y : \Omega \to E$ be a r.v with E countable. We want to define the conditional expectation of X given Y. From before, we have $\mathbb{E}[X \mid Y = y] = \dfrac{\mathbb{E}[X \mathbb{1}_{Y=y}]}{\mathbb{P}(Y=y)}$ for every y such that $\mathbb{P}(Y=y) > 0$.

Thus, we naturally set: $\mathbb{E}[X \mid Y] = \varphi(Y)$ where $\varphi : E \to \mathbb{R}$ is defined by $\varphi(y) = \begin{cases} \mathbb{E}[X \mid Y=y] & \text{if } \mathbb{P}(Y=y) > 0 \\ 0 & \text{otherwise} \end{cases}$

In other words $\mathbb{E}[X \mid Y]$ is a random variable defined by

$$\mathbb{E}[X \mid Y](\omega) = \varphi(Y(\omega)).$$

⚠ the notation $\mathbb{E}[X(\omega) \mid Y(\omega)]$ makes no sense.

The value of $\varphi(y)$ when $\mathbb{P}(Y=y)=0$ is arbitrary: it influences the definition of $\mathbb{E}[X|Y]$ only on a 0 probability set, the set $\{\omega\in\Omega: Y(\omega)\in E'\}$ with $E'=\{y\in E: \mathbb{P}(Y=y)=0\}$.

More generally, conditional expectations will always be defined up to 0 probability sets

Observe that $\mathbb{E}[X|Y]$ is a r.v which is $\sigma(Y)$ measurable, since it is a function of Y

END OF LECTURE 14

Example (dice toss) Take $\Omega=\{1,\dots,6\}$ with $\mathbb{P}(\{\omega\})=\frac{1}{6}$ for $\omega\in\Omega$. Set $X(\omega)=\omega$ and $Y(\omega)=\begin{cases}1 & \text{if } \omega \text{ odd}\\ 0 & \text{if } \omega \text{ even}\end{cases}$

Then $\mathbb{E}[X|Y](\omega)=\begin{cases}3 & \text{if } \omega\in\{1,3,5\}\\ 4 & \text{if } \omega\in\{2,4,6\}\end{cases}$

Lemma We have:

1) $\mathbb{E}[|\mathbb{E}[X|Y]|]\leq \mathbb{E}[|X|]$, so $\mathbb{E}[X|Y]\in L^1$

2) $\forall$ r.v Z which is $\sigma(Y)$-measurable and bounded, $\mathbb{E}[ZX]=\mathbb{E}[Z\,\mathbb{E}[X|Y]]$

(we take Z bounded to ensure integrability)

Proof 1) We have $\mathbb{E}[|\mathbb{E}[X|Y]|]=\sum_{y\in E}\mathbb{P}(Y=y)|\mathbb{E}[X|Y=y]|=\sum_{\substack{y\in E\\ \mathbb{P}(y)>0}}\mathbb{P}(y)\frac{|\mathbb{E}[X\mathbb{1}_{Y=y}]|}{\mathbb{P}(y)}\leq\sum_{\substack{y\in E\\ \mathbb{P}(y)>0}}\mathbb{E}[|X|\mathbb{1}_{Y=y}]$
$=\mathbb{E}[|X|]$

2) By the Doob-Dynkin lemma we can write $Z=F(Y)$. Then similarly,

$$\mathbb{E}[F(Y)\mathbb{E}[X|Y]]=\sum_{\substack{y\in E\\ \mathbb{P}(y)>0}}\mathbb{P}(Y=y)F(y)\mathbb{E}[X|Y=y]=\sum_{\substack{y\in E\\ \mathbb{P}(y)>0}}\mathbb{P}(Y=y)F(y)\frac{\mathbb{E}[X\mathbb{1}_{Y=y}]}{\mathbb{P}(Y=y)}$$

Which is $\mathbb{E}[X\sum_{\substack{y\in E\\ \mathbb{P}(y)>0}}F(y)\mathbb{1}_{Y=y}]=\mathbb{E}[XF(Y)]$ because a.s $\sum_{\substack{y\in E\\ \mathbb{P}(Y=y)>0}}F(y)\mathbb{1}_{Y=y}=F(Y)$.

∞

It turns out that this property can be used to define conditional expectations in a general setting

2) Definition and first properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $\mathcal{A} \subset \mathcal{F}$ is a σ-field, we write $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ if

- $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable
- $\mathbb{E}[|X|] < \infty$

Theorem Fix $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{A} \subset \mathcal{F}$ be a σ-field. Then there exists a random variable X' with

① $X' \in L^1(\Omega, \mathcal{A}, \mathbb{P})$

② For every r.v Z, $\mathcal{A}$ measurable and bounded, $\mathbb{E}[ZX] = \mathbb{E}[ZX']$

Moreover, if X'' is another such random variable, then $X' = X''$ a.s.

We denote by $\mathbb{E}[X \mid \mathcal{A}]$ any such random variable, called a version of the conditional expectation of X given $\mathcal{A}$.

Property ② is called the "characteristic property of conditional expectation"

We make some observations before the proof (assume for the moment that the theorem is true)

Remarks. $\mathbb{E}[X \mid \mathcal{A}]$ is an $\mathcal{A}$-measurable random, uniquely defined up to zero probability events. In practice, this is transparent because we consider only expectations of $\mathbb{E}[X \mid \mathcal{A}]$ or its almost sure properties. For this reason, we often say that "$\mathbb{E}[X \mid \mathcal{A}]$ is the conditional expectation of X given $\mathcal{A}$"

- By interpreting $\mathbb{E}[Z(X - X')] = 0$ by $\langle Z, X - \mathbb{E}[X|\mathcal{A}] \rangle = 0$, we can interpret $\mathbb{E}[X|\mathcal{A}]$ as the "projection" of X onto $\mathcal{A}$-measurable functions. This will be made rigorous when $X \in L^2$

Notation
- For $B \in \mathcal{F}$, we define $\mathbb{P}(B|\mathcal{A}) = \mathbb{E}[\mathbb{1}_B \mid \mathcal{A}]$: it is an $\mathcal{A}$-measurable r.v (defined a.s.)
- If $Y : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ is a r.v, we define $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)]$

(here X is always $\mathbb{R}$-valued integrable, but Y is not necessarily real-valued)

Remark

Recall the Doob-Dynkin lemma: If Y is $\mathbb{R}^n$-valued, then a $\sigma(Y)$ measurable function is of the form $f(Y)$ with f measurable. (we did the proof for $n=1$, but the proof is very similar). As a consequence:

- there is a measurable function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}[X \mid Y] = \varphi(Y)$
- If $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a measurable function such that $\mathbb{E}[|\varphi(Y)|] < \infty$ and for every f measurable bounded $\mathbb{E}[X f(Y)] = \mathbb{E}[\varphi(Y) f(Y)]$, then $\mathbb{E}[X \mid Y] = \varphi(Y)$

Remark the definition is consistent with what we saw in the discrete case. For example, is U is $\mathbb{Z}$-valued, let's find $\mathbb{E}[X \mid U]$. By the Doob-Dynkin lemma, $\mathbb{E}[X \mid U] = f(U)$ and we want to find f.
Taking $Z = \mathbb{1}_{U=n}$ gives $\mathbb{E}[X \mathbb{1}_{U=n}] = \mathbb{E}[f(U) \mathbb{1}_{U=n}] = \mathbb{E}[f(n) \mathbb{1}_{U=n}]$

Thus $f(n) = \dfrac{\mathbb{E}[X \mathbb{1}_{U=n}]}{\mathbb{P}(U=n)}$ when $\mathbb{P}(U=n) > 0$.

Simple properties of conditional expectation Take $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{A} \subset \mathcal{F}$ be a σ-field.

① $\mathbb{E}[\mathbb{E}[X \mid \mathcal{A}]] = \mathbb{E}[X]$ (very very useful)

② $\mathbb{E}[X \mid \{\emptyset, \Omega\}] = \mathbb{E}[X]$

③ If X is $\mathcal{A}$-measurable, $\mathbb{E}[X \mid \mathcal{A}] = X$. In particular, we always have $\mathbb{E}[X \mid \mathcal{F}] = X$.

④ $X \longmapsto \mathbb{E}[X \mid \mathcal{A}]$ is linear

⑤ $X \geq 0 \Rightarrow \mathbb{E}[X \mid \mathcal{A}] \geq 0$. As a consequence, $X_1 \geq X_2 \Rightarrow \mathbb{E}[X_1 \mid \mathcal{A}] \geq \mathbb{E}[X_2 \mid \mathcal{A}]$.

⑥ $|\mathbb{E}[X \mid \mathcal{A}]| \leq \mathbb{E}[|X| \mid \mathcal{A}]$

NB: In every statement, the "almost sure" is implicit (recall that conditional expectations are defined uniquely almost surely)

Proof ① Just take $Z = \mathbb{1}_\Omega$ in the characteristic property

② A $\{\emptyset, \Omega\}$-measurable function is constant, so $\mathbb{E}[X \mid \{\emptyset, \Omega\}]$ is constant, and thus equal to its expectation, which by ① is $\mathbb{E}[X]$.

③ Follows from uniqueness: X then satisfies ① and ②

④ If $X_1, X_2 \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, $x = \alpha \mathbb{E}[X_1 \mid \mathcal{A}] + \beta \mathbb{E}[X_2 \mid \mathcal{A}]$ satisfies for every $Z : \Omega \to \mathbb{R}$, $\mathcal{A}$-measurable, bounded, $\mathbb{E}[XZ] = \mathbb{E}[(\alpha X_1 + \beta X_2)Z]$.

so $\mathbb{E}[a X_1 + b X_2 | \mathcal{A}] = a\mathbb{E}[X_1|\mathcal{A}] + b\,\mathbb{E}[X_2|\mathcal{A}]$

5) Fix $\varepsilon > 0$. Take $Z = \mathbb{1}_{\mathbb{E}[X|\mathcal{A}] \leq -\varepsilon}$, $\mathcal{A}$-measurable. Then

$$0 \leq \mathbb{E}[X \mathbb{1}_{\mathbb{E}[X|\mathcal{A}] \leq -\varepsilon}] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}] \mathbb{1}_{\mathbb{E}[X|\mathcal{A}] \leq -\varepsilon}] \leq -\varepsilon\, \mathbb{P}(\mathbb{E}[X|\mathcal{A}] \leq -\varepsilon)$$

Hence $\mathbb{P}(\mathbb{E}[X|\mathcal{A}] \leq -\varepsilon) = 0$, so a.s $\mathbb{E}[X|\mathcal{A}] > -\varepsilon$.

Thus $\forall n \geq 1$, a.s. $\mathbb{E}[X|\mathcal{A}] > -\frac{1}{n}$. By interchanging "a.s" and "$\forall \in$ countable set":

a.s $\forall n \geq 1$, $\mathbb{E}[X|\mathcal{A}] > -\frac{1}{n}$. Thus a.s $\mathbb{E}[X|\mathcal{A}] \geq 0$.

6) Write $X = X^+ - X^-$ with $X^+, X^- \geq 0$. Then by 1)

$$|\mathbb{E}[X|\mathcal{A}]| = |\mathbb{E}[X^+|\mathcal{A}] - \mathbb{E}[X^-|\mathcal{A}]| \leq \mathbb{E}[X^+|\mathcal{A}] + \mathbb{E}[X^-|\mathcal{A}] = \mathbb{E}[X^+ + X^-|\mathcal{A}] = \mathbb{E}[|X||\mathcal{A}]$$

END OF LECTURE 15

Proof of the theorem

Uniqueness: Assume that X' and X'' satisfy 1) and 2) above. Take $Z = \mathbb{1}_{\{X' > X''\}}$, $\mathcal{A}$-measurable.

Then $\mathbb{E}[(X' - X'')\mathbb{1}_{X' > X''}] = \mathbb{E}[X' \mathbb{1}_{X' > X''}] - \mathbb{E}[X'' \mathbb{1}_{X' > X''}] = \mathbb{E}[X \mathbb{1}_{X' > X''}] - \mathbb{E}[X \mathbb{1}_{X' > X''}] = 0$

But a.s $(X' - X'')\mathbb{1}_{X' > X''} \geq 0$. Thus a.s $(X' - X'')\mathbb{1}_{X' > X''} = 0$. Thus $\mathbb{P}(X' > X'') = 0$, so a.s $X' \leq X''$. By symmetry, $X'' \leq X'$ a.s, so $X' = X''$ a.s.

existence: There are essentially two approaches using results from measure theory: one with Radon-Nikodym, one with Hilbert L^2 spaces. We present the second approach and underline in blue facts from measure theory that we do not prove.

Step 1: Assume $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. We can equip $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with a scalar product: for $Y, Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, we set $\langle Y Z \rangle = \mathbb{E}[YZ]$, so that the norm $\|\cdot\|$ defined by $\|Y\|^2 = \langle Y|Y \rangle$ defines a normed vector space (here we identify two r.v which are equal a.s.), which is complete: $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space. Then $L^2(\Omega, \mathcal{A}, \mathbb{P})$ is a closed subset of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ so we can consider the orthogonal projection X' of X on $L^2(\Omega, \mathcal{A}, \mathbb{P})$ characterized by the fact that $\forall Z \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, $\langle X - X' | Z \rangle = 0$, which implies

$\mathbb{E}[XZ] = \mathbb{E}[X'Z]$, and gives the result (since Z $\mathcal{A}$-measurable, bounded implies $Z \in L^2(\Omega, \mathcal{A}, \mathbb{P})$)

Step 2: Assume $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $X \geq 0$. We use a truncation argument: for $n \geq 1$, set $X_n = X \mathbb{1}_{X \leq n} \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. Set $X'_n = \mathbb{E}[X_n | \mathcal{A}]$. Since $X_n \leq X_{n+1}$ we have $X'_n \leq X'_{n+1}$ a.s by property ③ above. By monotonicity, we can define $X' = \lim_{n \to \infty} X'_n$ as an a.s. limit. Then for $Z \geq 0$ $\mathcal{A}$-measurable bounded, from $\mathbb{E}[X'_n Z] = \mathbb{E}[X_n Z]$ we get $\mathbb{E}[X' Z] \overset{(*)}{=} \mathbb{E}[XZ]$ by monotone convergence. When Z is not necessarily ≥ 0, we write $Z = Z^+ - Z^-$ and apply $(*)$ with Z^+ and Z^- and get $(*)$ with Z by linearity.

By taking $Z = 1$ we get $\mathbb{E}[X'] = \mathbb{E}[X] < \infty$.

Step 3 When $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, write $X = X^+ - X^-$. We then set

$$X' = \mathbb{E}[X^+ | \mathcal{A}] - \mathbb{E}[X^- | \mathcal{A}],$$ which satisfies the requirements ① and ②

3) Conditional expectation of non-negative r.v.

We can actually define conditional expectations for r.v in $[0, \infty]$ without integrability conditions

Theorem Fix a r.v X $[0,\infty]$-valued. Let $\mathcal{A} \subset \mathcal{F}$ be a σ-field. Then there exists a random variable X' with

① $X' \in [0, \infty]$, X' is $\mathcal{A}$ measurable

② For every r.v Z, $\mathcal{A}$ measurable and ≥ 0, $\mathbb{E}[ZX] = \mathbb{E}[ZX']$

Moreover, if X'' is another such random variable, then $X' = X''$ a.s.

We denote by $\mathbb{E}[X | \mathcal{A}]$ any such random variable, called a version of the conditional expectation of X given $\mathcal{A}$.

Proof: Existence As in step 2 above, if we define $X_n = X \mathbb{1}_{X \leq n}$ and $X'_n = \mathbb{E}[X_n | \mathcal{A}]$, X', the a.s limit of $(X'_n)_{n \geq 1}$ satisfies ① and ②

Uniqueness Assume that X' and X'' satisfy ① and ②. Then for $a < b \in \mathbb{Q}_+$, by taking $Z = \mathbb{1}_{\{X' \leq a < b \leq X''\}}$ we get

$$b\, \mathbb{P}(X' \leq a < b \leq X'') = \mathbb{E}[bZ] \leq \mathbb{E}[X'' Z] = \mathbb{E}[X' Z] \leq a\, \mathbb{P}(X' \leq a < b \leq X'')$$

Since $a < b$, we get $\mathbb{P}(X' \leq a < b \leq X'') = 0$. By taking a countable union we get $\mathbb{P}(X' < X'') = 0$. thus $\mathbb{P}(X' \geq X'') = 1$. By symmetry $\mathbb{P}(X' \leq X'') = 1$, so $\mathbb{P}(X' = X'') = 1$

As for conditional expectations in L^1, we have the following properties which immediately follow from the definition:

Simple properties of conditional expectation Take X $[0,\infty]$ valued r.v and let $\mathcal{A} \subset \mathcal{F}$ be a σ-field.

1. $\mathbb{E}[X \mid \{\phi, \Omega\}] = \mathbb{E}[X]$, $\mathbb{E}[X \mid \mathcal{F}] = X$
2. If X is $\mathcal{A}$-measurable, $\mathbb{E}[X \mid \mathcal{A}] = X$.
3. if $Y \in [0,\infty]$, $a, b \geq 0$, $\mathbb{E}[aX + bY \mid \mathcal{A}] = a\mathbb{E}[X \mid \mathcal{A}] + b\,\mathbb{E}[Y \mid \mathcal{A}]$
4. $\mathbb{E}[\mathbb{E}[X \mid \mathcal{A}]] = \mathbb{E}[X]$ (Very very useful)
5. $X_1 \geq X_2 \geq 0 \Rightarrow \mathbb{E}[X_1 \mid \mathcal{A}] \geq \mathbb{E}[X_2 \mid \mathcal{A}]$.

4) Convergence theorems

Theorem Let $\mathcal{A} \subset \mathcal{F}$ be a σ-field

1. [Conditional monotone convergence] Let $(X_n)_{n \geq 0}$ be an increasing sequence of $[0,\infty]$ r.v with $X = \lim_{n\to\infty} \nearrow X_n$

   Then $\mathbb{E}[X_n \mid \mathcal{A}] \underset{n\to\infty}{\nearrow} \mathbb{E}[X \mid \mathcal{A}]$
2. [Conditional Fatou] Let $(X_n)_{n \geq 0}$ be a sequence of $[0,\infty]$ r.v. Then $\mathbb{E}[\liminf_{n\to\infty} X_n \mid \mathcal{A}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{A}]$
3. [Conditional dominated convergence] Let $(X_n)_{n \geq 1}$ be a sequence of integrable r.v. such that
   - $X_n \xrightarrow{a.s} X$
   - $\exists Y \geq 0$ in L^1 s.t $\forall n \geq 1$, a.s. $|X_n| \leq Y$

   Then $\mathbb{E}[X_n \mid \mathcal{A}] \xrightarrow[n\to\infty]{} \mathbb{E}[X \mid \mathcal{A}]$ almost surely and in $L^1(\Omega, \mathcal{A}, \mathbb{P})$
4. [Conditional Jensen] Let $f : \mathbb{R} \to \mathbb{R}_+$ be a convex function (if $f'' \geq 0$, f is convex) such that $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$

   Then $f(\mathbb{E}[X \mid \mathcal{A}]) \leq \mathbb{E}[f(X) \mid \mathcal{A}]$

By taking $\mathcal{A} = \{\phi, \Omega\}$ we get $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$ (Jensen inequality)

(as usual for conditional expectations, all statements are to be understood almost surely)

Proof ① We know that $\mathbb{E}[X_n|\mathcal{A}]$ is $\nearrow$. Let X' be its limit, $\mathcal{A}$ measurable as a limit of $\mathcal{A}$-measurable functions. Then for $Z\geq 0$ $\mathcal{A}$ measurable, $\mathbb{E}[X_n Z] = \mathbb{E}[\mathbb{E}[X_n|\mathcal{A}] Z]$, so by monotone convergence we get $\mathbb{E}[XZ]=\mathbb{E}[X'Z]$ Thus $X' = \mathbb{E}[X|\mathcal{A}]$ a.s

② Observe that for $i\geq n\geq 1$

$$\mathbb{E}[\inf_{k\geq n} X_k|\mathcal{A}] \leq \mathbb{E}[X_i|\mathcal{A}].$$

Thus $\mathbb{E}[\inf_{k\geq n} X_k|\mathcal{A}] \leq \inf_{k\geq n} \mathbb{E}[X_k|\mathcal{A}]$.

But $\inf_{k\geq n} X_k \underset{n\to\infty}{\nearrow} \liminf_{n\to\infty} X_k$, so the result follows from ①

③ Apply ② with $Z-X_n\geq 0$ and $Z+X_n\geq 0$.

$$\mathbb{E}[Z|\mathcal{A}]-\mathbb{E}[X|\mathcal{A}] = \mathbb{E}[\liminf_{n\to\infty}(Z-X_n)|\mathcal{A}] \leq \mathbb{E}[Z|\mathcal{A}] - \limsup_{n\to\infty}\mathbb{E}[X_n|\mathcal{A}]$$

$$\mathbb{E}[Z|\mathcal{A}]+\mathbb{E}[X|\mathcal{A}] = \mathbb{E}[\liminf_{n\to\infty}(Z+X_n)|\mathcal{A}] \leq \mathbb{E}[Z|\mathcal{A}] + \liminf_{n\to\infty}\mathbb{E}[X_n|\mathcal{A}]$$

Thus $\mathbb{E}[X|\mathcal{A}] \leq \liminf_{n\to\infty}\mathbb{E}[X_n|\mathcal{A}] \leq \limsup_{n\to\infty}\mathbb{E}[X_n|\mathcal{A}] \leq \mathbb{E}[X|\mathcal{A}]$ Which gives the a.s convergence

The L^1 convergence follows from dominated convergence, since $|\mathbb{E}[X_n|\mathcal{A}]| \leq \mathbb{E}[|X_n||\mathcal{A}] \leq \mathbb{E}[Y|\mathcal{A}]$ which is in L^1 since $\mathbb{E}[\mathbb{E}[Y|\mathcal{A}]] = \mathbb{E}[Y]<\infty$

④ Set $E_f = \{(a,b)\in\mathbb{R}^2 : \forall x\in\mathbb{R},\ f(x)\geq ax+b\}$. Since f is convex, $\forall x\in\mathbb{R},\ f(x) = \sup_{(a,b)\in E_f}(ax+b)$ (*). In addition, we can find $D\subset E_f$ countable and dense in E_f (**)

Thus $\mathbb{E}[f(X)|\mathcal{A}] = \mathbb{E}[\sup_{(a,b)\in D}(aX+b)|\mathcal{A}] \geq \sup_{(a,b)\in D} \overbrace{\mathbb{E}[aX+b|\mathcal{A}]}^{= a\mathbb{E}[X|\mathcal{A}]+b} = f(\mathbb{E}[X|\mathcal{A}])$

(here we need countability because conditional expectations are only defined a.s)

Justification of (*) For $x\in\mathbb{R}$ we clearly have $f(x)\geq \sup_{(a,b)\in E_f}(ax+b)$. The other inequality comes from the fact that by convexity of f we can find $a,b\in\mathbb{R}$ such that $f(x)=ax+b$ and $f(y)\geq ay+b$ $\forall y\in\mathbb{R}$.

Justification of (**) Denote by $\{r_n : n\geq 1\}$ the elements of $\mathbb{Q}^2$. Let $p_n\in E_f$ be such that $|p_n - r_n| \leq d(E_f, r_n)+\frac{1}{n}$ where $d(E_f,r_n) = \inf_{p\in E}|p-r_n|$. Let us check that $\{p_n : n\geq 1\}$ is dense in E_f. Fix $\varepsilon>0$. By density, we can find $n\geq\frac{1}{\varepsilon}$ with $|r_n - x|\leq\varepsilon$

then $|p_n - x| \le |p_n - r_n| + |r_n - x| \le \frac{1}{n} + d(r_n, E_\delta) + \varepsilon \le 2\varepsilon + d(r_n, E_\delta)$.
But $d(r_n, E_\delta) \le |r_n - x| \le \varepsilon$ (because $x \in E_\delta$), so $|p_n - x| \le 3\varepsilon$.

Remark We often use Jensen's inequality with $f(x) = |x|^p$ with $p \ge 1$ and with the r.v $|X|$:
$\mathbb{E}[|X|]^p \le \mathbb{E}[|X|^p]$ and $\mathbb{E}[|X| \mid \mathcal{A}]^p \le \mathbb{E}[|X|^p \mid \mathcal{A}]$

5) Some useful properties

The "information contained in $\mathcal{A}$" can be factorized out of the conditional expectation. From now on by L^1 we mean $L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Proposition Let $\mathcal{A} \subset \mathcal{F}$ be a σ-field. Let X, Y be real-valued r.v such that $X, Y \in [0,\infty]$ or $X, XY \in L^1$.
Assume that Y is $\mathcal{A}$-measurable. Then $\mathbb{E}[YX \mid \mathcal{A}] = Y\,\mathbb{E}[X \mid \mathcal{A}]$

Proof If $X, Y \in [0,\infty]$ and $Z \ge 0$ is $\mathcal{A}$-measurable, then $ZY \ge 0$ is $\mathcal{A}$ measurable, so
$\mathbb{E}[ZYX] = \mathbb{E}[ZY\,\mathbb{E}[X \mid \mathcal{A}]]$
$Y\,\mathbb{E}[X \mid \mathcal{A}]$ is $\mathcal{A}$-measurable and satisfies the characteristic property of conditional expectation, so $Y\,\mathbb{E}[X \mid \mathcal{A}] = \mathbb{E}[YX \mid \mathcal{A}]$. The proof is similar when $X, XY \in L^1$

One has the tower property (restricting information)

END OF LECTURE 16

Proposition Let $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{F}$ be σ-fields. Then for X r.v. with $X \in [0,\infty]$ or $X \in L^1$,
$\mathbb{E}[\,\mathbb{E}[X \mid \mathcal{A}_2] \mid \mathcal{A}_1] = \mathbb{E}[X \mid \mathcal{A}_1]$

Proof Let $Z \ge 0$ be $\mathcal{A}_1$ measurable bounded. By uniqueness it is enough to check that
$\mathbb{E}[ZX] = \mathbb{E}[Z\,\mathbb{E}[\mathbb{E}[X \mid \mathcal{A}_2] \mid \mathcal{A}_1]]$
To this end, write:
$$\mathbb{E}[Z\,\mathbb{E}[\mathbb{E}[X \mid \mathcal{A}_2] \mid \mathcal{A}_1]] = \mathbb{E}[Z\,\mathbb{E}[X \mid \mathcal{A}_2]] = \mathbb{E}[ZX]$$
because Z is also $\mathcal{A}_2$ measurable. The proof is similar for $X \in L^1$.

Adding independent information does not change the conditional expectation:

Lemma Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{F}$ be σ-fields and X a r.v which is $\in [0,\infty]$ or in L^1. If $\mathcal{A}_2 \perp\!\!\!\perp \sigma(\sigma(X), \mathcal{A}_1)$, then $\mathbb{E}[X \mid \sigma(\mathcal{A}_1, \mathcal{A}_2)] = \mathbb{E}[X \mid \mathcal{A}_1]$

Proof: We show that $\mathbb{E}[\mathbb{1}_C X] \overset{(*)}{=} \mathbb{E}[\mathbb{1}_C \mathbb{E}[X \mid \sigma(\mathcal{A}_1)]]$ for every C in a generating π-system of $\sigma(\mathcal{A}_1, \mathcal{A}_2)$. Indeed, since $\mathbb{E}[X \mid \sigma(\mathcal{A}_1)]$ is $\sigma(\mathcal{A}_1, \mathcal{A}_2)$ measurable, this will imply the result (see Exercise 2 Sheet 8).

We use $\{A_1 \cap A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$, which is a generating π-system of $\sigma(\mathcal{A}_1, \mathcal{A}_2)$.

For $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$,

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{A_1 \cap A_2} X] &= \mathbb{E}[\mathbb{1}_{A_2} \mathbb{1}_{A_1} X] \\ &= \mathbb{E}[\mathbb{1}_{A_2}]\, \mathbb{E}[\mathbb{1}_{A_1} X] \quad \text{by } \perp\!\!\!\perp \\ &= \mathbb{E}[\mathbb{1}_{A_2}]\, \mathbb{E}[\mathbb{1}_{A_1} \mathbb{E}[X \mid \mathcal{A}_1]] \\ &= \mathbb{E}[\mathbb{1}_{A_2} \cdot \mathbb{1}_{A_1} \mathbb{E}[X \mid \mathcal{A}_1]] \end{aligned}$$

and we get $(*)$ $\sim$

Corollary If $X \perp\!\!\!\perp Y$, $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$

(take $\mathcal{A}_1 = \{\emptyset, \Omega\}$ and $\mathcal{A}_2 = \sigma(Y)$)

b) Conditional density functions **(not covered in class: optional)**

Assume that X, Y take values in $\mathbb{R}^m, \mathbb{R}^n$ and that (X,Y) has a density: $\mathbb{P}_{(X,Y)}(dx\,dy) = f_{(X,Y)}(x,y)\,dx\,dy$. Let $f_Y(y) = \int_{\mathbb{R}^m} f_{(X,Y)}(x,y)\,dx$ be a density of Y.

Then for $h : \mathbb{R}^m \to \mathbb{R}_+$ and $g : \mathbb{R}^n \to \mathbb{R}_+$ measurable, we have

$$\mathbb{E}[h(X) g(Y)] = \int_{\mathbb{R}^m \times \mathbb{R}^n} h(x) g(y) f_{(X,Y)}(x,y)\,dx\,dy \quad \text{(transfer theorem)}$$

$$= \int_{\mathbb{R}^n} g(y) f_Y(y)\,dy \int_{\mathbb{R}^m} \frac{h(x) f_{(X,Y)}(x,y)}{f_Y(y)} \mathbb{1}_{f_Y(y)>0}\,dx$$

$$= \mathbb{E}[\varphi(Y) g(Y)]$$

With $\varphi(y) = \begin{cases} \dfrac{1}{f_Y(y)} \displaystyle\int_{\mathbb{R}^m} h(x)\, f_{(X,Y)}(x,y)\, dx & \text{if } f_Y(y)>0 \\ 0 & \text{otherwise} \end{cases}$

Thus $\mathbb{E}[h(X) \mid Y] = \varphi(Y)$.

We interpret this result by writing

$$\mathbb{E}[h(X) \mid Y] = \int_{\mathbb{R}^m} h(x)\, \nu(Y, dx)$$

where $\nu(y, dx) = \dfrac{1}{f_Y(y)} f_{(X,Y)} \mathbb{1}_{f_Y(y)>0}\, dx = f_{X|Y}(x|y)\, dx$. The measure $\nu(y, dx)$ is called <u>conditional distribution given $Y=y$</u> and $f_{X|Y}(x|y)$ is the <u>conditional density function of X given $Y=y$</u>. Notice that this function is defined only up to a zero-measure set.