Lecture 4 · 23.4.2021 · 10:15–12:00 via Zoom

Uniform integrability, conditional expectation.

1 Uniform integrability (continuation)

Definition 1. *A family $\mathcal{X}$ of random variables is uniformly integrable (UI) if for any $\varepsilon > 0$ there exists $L > 0$ such that*

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

In particular, it holds that

$$\sup_{X \in \mathcal{X}} \mathbb{E}[|X|] < \infty .$$

Uniform integrability says that large values of the r.v.s. contribute uniformly little to the averages.

Example 2. A single integrable r.v. X is uniformly integrable. (Exercise: use monotone convergence). A finite family of integrable r.v. is also uniformly integrable (Exercise: also easy)

Example 3. Let U be a uniform r.v. on $[0, 1]$. Take $X_n := n \mathbb{1}_{U \leqslant 1 / n}$, then the family $(X_n)_{n \geqslant 1}$ is not UI. Indeed for every $L \geqslant 1$ one has

$$\mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}] = n \mathbb{E}[\mathbb{1}_{U \leqslant 1 / n} \mathbb{1}_{n \geqslant L}] = \mathbb{1}_{n \geqslant L} n \mathbb{P}(U \leqslant 1 / n) = \mathbb{1}_{n \geqslant L}$$

this means that for any $L \geqslant 1$

$$\sup_{n \geqslant 1} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}] = 1$$

so we cannot make it (uniformly in n) small by choosing L large. The family is not UI. For this faamily we have $X_n \rightarrow 0$ a.s. but $\mathbb{E}[X_n] = 1$ for all $n \geqslant 1$.

Let's give now an equivalent reformulation of UI:

Lemma 4. *A family $\mathcal{X}$ of r.v.s is UI iff*

1. *The family is bounded in L^1, i.e.*

$$\sup_{X \in \mathcal{X}} \mathbb{E}[|X|] < \infty ,$$

2. *For all $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that for all $A \in \mathcal{F}$ for which $\mathbb{P}(A) \leqslant \delta$ we have*

$$\sup_{X \in \mathcal{X}} \mathbb{E}[|X| \mathbb{1}_A] \leqslant \varepsilon .$$

Proof. ($\Leftarrow$) Let's assume properties 1. and 2. and prove UI. Property 2. tells me that for every $\varepsilon > 0$, we have $\delta > 0$ so that $\sup_{X \in \mathscr{X}} \mathbb{E}[|X| \mathbb{1}_A] \leqslant \varepsilon$ whenever $\mathbb{P}(A) \leqslant \delta$. We want to apply this to $A = \{|X| \geqslant L\}$ for some L. Note that

$$\mathbb{P}(|X| \geqslant L) \underset{\text{Markov}}{\leqslant} \frac{\mathbb{E}[|X|]}{L} \leqslant L^{-1} \sup_{Y \in \mathscr{X}} \mathbb{E}[|Y|] \leqslant C L^{-1}.$$

I want this to be $\leqslant \delta$ so I choose $L = L(\varepsilon) \geqslant C \delta(\varepsilon)^{-1}$. For such L we have $\sup_{X \in \mathscr{X}} \mathbb{E}[|X| \mathbb{1}_{|X| \geqslant L}] \leqslant \varepsilon$ therefore UI holds.

($\Rightarrow$) Note that $|X| \leqslant K + (|X| - K)_+$ for any $K \geqslant 0$. So we have

$$\mathbb{E}[|X| \mathbb{1}_A] \leqslant \mathbb{E}[K \mathbb{1}_A] + \mathbb{E}[(|X| - K)_+ \mathbb{1}_A] \leqslant K\, \mathbb{P}(A) + \mathbb{E}[(|X| - K)_+]$$

$$\leqslant K\, \mathbb{P}(A) + \mathbb{E}[|X| \mathbb{1}_{|X| \geqslant K}]$$

so by UI of the family $\mathscr{X}$ we can choose $K = K(\varepsilon)$ large so that $\mathbb{E}[|X| \mathbb{1}_{|X| \geqslant K}] \leqslant \varepsilon / 2$ for all $X \in \mathscr{X}$ and then I can choose $\delta = \delta(\varepsilon)$ small enough so that $K(\varepsilon) \delta(\varepsilon) \leqslant \varepsilon / 2$. With these choices I have

$$\sup_{X \in \mathscr{X}} \mathbb{E}[|X| \mathbb{1}_A] \leqslant K(\varepsilon)\, \mathbb{P}(A) + \frac{\varepsilon}{2} \leqslant \varepsilon$$

for every $A \in \mathscr{F}$ such that $\mathbb{P}(A) \leqslant \delta(\varepsilon)$. This proves prop. 2. and we have already seen that UI implies uniform bound on the L^1 of the family (was an exercise). □

UI is important because is the best condition for convergence of integrals:

Theorem 5. (Uniform integrability) *Let $(X_n)_n$ and X be integrable random variables, then*

$$\mathbb{E}[|X_n - X|] \to 0$$

(convergence in average or in L^1) iff

a) $X_n \to X$ in probability, i.e. $\lim_n \mathbb{P}(|X_n - X| > \varepsilon) = 0$ for all $\varepsilon > 0$;

b) the family $(X_n)_n$ is uniformly integrable.

Remark 6. Note that convergence in L^1 implies convergence of expectations:

$$|\mathbb{E}[X_n] - \mathbb{E}[X]| \leqslant \mathbb{E}[|X_n - X|] \to 0$$

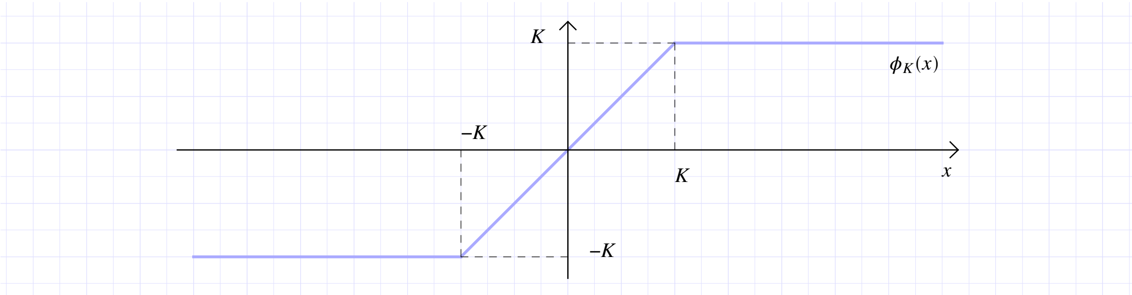
if $X_n \to X$ in L^1.

Proof. ($\Leftarrow$) Note that conv. in probability is enough to prove L^1 convergence if the r.v. are uniformly bounded in L^∞ (that is they are pointwise bounded a.s.). The large values of the sequence are taken care by uniform integrability. Let's write down the details. Consider the function

$$\phi_K(x) = ((K \wedge x) \vee (-K)), \qquad x \in \mathbb{R}$$

which is such that $|\phi_K(x)-\phi_K(y)| \leqslant [|x-y| \wedge (2K)]$ for all $x, y \in \mathbb{R}$ and

$$|\phi_K(x) - x| \leqslant (|x| - K)_+ \leqslant |x| \, \mathbb{1}_{|x| \geqslant K}$$



Now we have for any K, δ

$$\mathbb{E}[|X_n - X|] \leqslant \mathbb{E}[|\phi_K(X_n) - \phi_K(X)|] + \mathbb{E}[|X_n - \phi_K(X_n)|] + \mathbb{E}[|X - \phi_K(X)|]$$

$$\leqslant \mathbb{E}\Big[\underbrace{|\phi_K(X_n) - \phi_K(X)|}_{\leqslant |X_n - X| \leqslant \delta} \mathbb{1}_{|X_n - X| \leqslant \delta}\Big] + \mathbb{E}\Big[\underbrace{|\phi_K(X_n) - \phi_K(X)|}_{\leqslant (2K)} \mathbb{1}_{|X_n - X| > \delta}\Big] + \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant K}] + \mathbb{E}[|X| \mathbb{1}_{|X| \geqslant K}]$$

$$\leqslant \delta + (2K) \mathbb{P}(|X_n - X| > \delta) + \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant K}] + \mathbb{E}[|X| \mathbb{1}_{|X| \geqslant K}]$$

so by convergence in prob. we have $\mathbb{P}(|X_n - X| > \delta) \to 0$ for any fixed δ while by UI we can choose K large enough so that $\mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant K}] + \mathbb{E}[|X| \mathbb{1}_{|X| \geqslant K}] \leqslant \varepsilon$. So we have

$$0 \leqslant \limsup_n \mathbb{E}[|X_n - X|] \leqslant \delta + \varepsilon$$

but since ε and δ are arbitrary we have proven $\mathbb{E}[|X_n - X|] \to 0$.

($\Rightarrow$) We have to prove now that convergence in L^1 implies both conv. in prob. and UI. For conv. in prob. it is easy: it suffice to remark that for every $\delta > 0$

$$\mathbb{P}(|X_n - X| > \delta) \leqslant \frac{\mathbb{E}[|X_n - X|]}{\delta} \to 0.$$

We have to prove now that $\mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}]$ can be made uniformly small by choosing L. We remark that pointwise in Ω we have $|X_n| \leqslant |X - X_n| + |X|$ therefore by integrating on the set $\{|X_n| \geqslant L\}$ we have

$$\mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}] = \mathbb{E}[|X_n - X| \mathbb{1}_{|X_n| \geqslant L}] + \mathbb{E}[|X| \mathbb{1}_{|X_n| \geqslant L}]$$

$$\leqslant \mathbb{E}[|X_n - X|] + \mathbb{E}[|X| \mathbb{1}_{|X_n| \geqslant L}]$$

and we use that X is a uniformly integrable random variable, so in particular by Lemma 4 for any $\varepsilon > 0$ we have $\delta(\varepsilon) > 0$ so that $\mathbb{E}[|X| \mathbb{1}_A] \leqslant \varepsilon / 3$ for all A such that $\mathbb{P}(A) \leqslant \delta(\varepsilon)$. Morever by Markov we have

$$\mathbb{P}(|X_n| \geqslant L) \leqslant L^{-1} \mathbb{E}[|X_n|] \leqslant L^{-1} \sup_m \mathbb{E}[|X_m|].$$

Note that $\sup_m \mathbb{E}[|X_m|] < \infty$ since the family $(X_m)_m$ converges in L^1 and therefore is bounded in this Banach space and I can choose $L_1(\varepsilon)$ so that for all $L \geqslant L_1(\varepsilon)$ we have $\mathbb{P}(|X_n| \geqslant L) \leqslant \delta(\varepsilon)$ and therefore that

$$\mathbb{E}[|X| \mathbb{1}_{|X_n| \geqslant L}] \leqslant \frac{\varepsilon}{3}, \qquad L \geqslant L_1(\varepsilon), n \geqslant 1,$$

Moreover by L^1 convergence we have that there exists $n_0(\varepsilon)$ such that for all $n \geqslant n_0(\varepsilon)$ we have $\mathbb{E}[|X_n - X|] \leqslant \varepsilon / 3$. Therefore we have now

$$\sup_{n \geqslant 1} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}] \leqslant \sup_{n:1 \leqslant n \leqslant n_0(\varepsilon)} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}] + \sup_{n:n \geqslant n_0(\varepsilon)} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}]$$

$$\leqslant \sup_{n:1 \leqslant n \leqslant n_0(\varepsilon)} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}] + \frac{2\varepsilon}{3}$$

and now the family $(X_n)_{n=1,\ldots,n_0(\varepsilon)}$ is finite so also UI and we can find $L_2(\varepsilon)$ so that

$$\sup_{n:1 \leqslant n \leqslant n_0(\varepsilon)} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L_2}] \leqslant \frac{\varepsilon}{3}$$

and taking $L(\varepsilon) = \max(L_1(\varepsilon), L_2(\varepsilon))$ we have that

$$\sup_{n \geqslant 1} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geqslant L}] \leqslant \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} = \varepsilon,$$

which proves UI of the family. □

Remark 7. In the exercise sheet there are further important properties of UI r.v. For example you will prove that:

- If a family is uniformly bounded in L^p for some $p > 1$ then it is also UI.
- If

  $$\sup_{X \in \mathscr{X}} \mathbb{E}[|g(X)|] < \infty$$

  for some function $g: \mathbb{R} \to \mathbb{R}_+$ increasing at finity faster than x then the family $\mathscr{X}$ is UI.
- If $\mathscr{X}, \mathscr{Y}$ are two UI families then $\mathscr{X} + \mathscr{Y} = \{X + Y : X \in \mathscr{X}, Y \in \mathscr{Y}\}$ is UI.

2 Conditional expectations

We introduce now the general notion of conditional expectation.

Fix a prob. space $(\Omega, \mathscr{F}, \mathbb{P})$.

Remember that given a set $B \in \mathscr{F}$ such that $\mathbb{P}(B) > 0$ we can introduce a new probability measure $\mathbb{P}_B: \mathscr{F} \to [0, 1]$ by letting

$$\mathbb{P}_B(A) := \mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

The measure $\mathbb{P}_B$ represent the probability of event *conditionally* on the occurrence of the event B. The symbol $\mathbb{P}(A|B)$ is the conditional probability of the even A given the event B.

We say that A is independent of B when $\mathbb{P}_B(A) = \mathbb{P}(A)$. This is equivalent to ask that

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

Independence between events depend on the probability measure. We say that two σ-algebras $\mathscr{G}$, $\mathscr{H}$ are independent iff

$$\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B), \qquad A\in\mathscr{G}, B\in\mathscr{H}.$$

Two random variables X, Y are independent when $\sigma(X)$ and $\sigma(Y)$ are independent.

All these notions extents to arbitrary families.

This definition of conditional expectation is ok for elementary applications but we run early into problems in more general applications.

Example 8. Let X, Y two uniform random variables on $[0,1]$ we cannot define interesting conditional probabilities like

$$\mathbb{P}(X\in[1/3,1/2]|Y=0.6)$$

since $\mathbb{P}(Y=0.6)=0$, however is clear from the intuition that we should be able to have such a quantity. Practically one could image to proceed as follows: take $\delta>0$ and compute

$$\mathbb{P}(X\in[1/3,1/2]|Y\in[0.6-\delta,0.6+\delta])=\frac{\mathbb{P}(X\in[1/3,1/2], Y\in[0.6-\delta,0.6+\delta])}{\mathbb{P}(Y\in[0.6-\delta,0.6+\delta])}$$

which is well defined, and then we try to take the limit $\delta\to0$. But this is an indeterminate form $0/0$ in general so we need more informations. In particular we can assume that (X,Y) has a joint density, so that

$$\mathbb{P}((X,Y)\in A)=\int_A f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$

for all $A\in\mathscr{B}(\mathbb{R}^2)$. In this case we have

$$\mathbb{P}(X\in[1/3,1/2]|Y\in[0.6-\delta,0.6+\delta])=\frac{\int_{x\in[1/3,1/2],y\in[0.6-\delta,0.6+\delta]}f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y}{\int_{y\in[0.6-\delta,0.6+\delta]}f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y}.$$

If we assume moreover that the density $f_{X,Y}\colon\mathbb{R}^2\to\mathbb{R}_+$ is a continuous function, one can show that the limit $\delta\to0$ exists and give

$$\lim_{\delta\to0}\frac{\int_{x\in[1/3,1/2],y\in[0.6-\delta,0.6+\delta]}f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y}{\int_{y\in[0.6-\delta,0.6+\delta]}f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y}=\frac{\int_{x\in[1/3,1/2]}f_{X,Y}(x,0.6)\mathrm{d}x}{\int f_{X,Y}(x,0.6)\mathrm{d}x}$$

provided of course that $\int f_{X,Y}(x,0.6)\mathrm{d}x>0$. So in this case we can define more generally

$$\mathbb{P}(X\in A|Y=y):=\frac{\int_{x\in A}f_{X,Y}(x,y)\mathrm{d}x}{\int f_{X,Y}(x,y)\mathrm{d}x}.$$

Which confirms our intuition that this conditional probability should in some sense exist.

This way of proceeding is quite coumbersome and depends on many details we would like not to care about in a general theory.

We need a more robust approach, in particular which does not depend on taking ratios. It should give back the above two definitions in the associated particular cases.

The first step is to avoid taking ratios. To do that we need to put attention on expectations more than on proabilities. Assume we have two random variables X, Y as above and we try to condition one on the values of the other.

Let's start assuming that they are both discrete r.v.. Note that for every measurable and bounded $f: \mathbb{R}^2 \to \mathbb{R}$ we have the formula

$$\mathbb{E}[f(X,Y)] = \sum_{x,y} f(x,y)\mathbb{P}(X=x, Y=y)$$

$$= \sum_y \underbrace{\sum_x f(x,y)\mathbb{P}(X=x|Y=y)}_{g(y)}\mathbb{P}(Y=y) = \sum_y g(y)\mathbb{P}(Y=y) = \mathbb{E}[g(Y)]$$

(provided we *define* $\mathbb{P}(X=x|Y=y)=0$ when $\mathbb{P}(Y=y)=0$). Here we have intoduced the function

$$g(y) := \begin{cases} \sum_x f(x,y)\mathbb{P}(X=x|Y=y) & \text{when } \mathbb{P}(Y=y) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Moreover note that if $f(x,y) = \mathbb{1}_{x \in A}$ then $g(y) = \mathbb{P}(X \in A|Y=y)$ when $\mathbb{P}(Y=y) \neq 0$ so we get back the elementary definition of conditional prob.

What is the general relation between the function f and the function g? We have that for *any* measurable and bounded $h: \mathbb{R} \to \mathbb{R}$ we also have

$$\mathbb{E}[f(X,Y)h(Y)] = \mathbb{E}[g(Y)h(Y)]. \tag{1}$$

This is a family of equations for g given f. It is easy to see that if all these equations are satisfied then in the discrete case we have indeed the function g as defined above. Indeed take $h(y) = \mathbb{1}_{y=z}$ then

$$g(z)\mathbb{P}(Y=z) = \mathbb{E}[g(Y)h(Y)] = \mathbb{E}[f(X,Y)h(Y)]$$

$$= \sum_{x,y} f(x,y)h(y)\mathbb{P}(X=x, Y=y) = \sum_x f(x,z)\mathbb{P}(X=x, Y=z)$$

so if $\mathbb{P}(Y=z) \neq 0$ we have

$$g(z) = \frac{\sum_x f(x,z)\mathbb{P}(X=x, Y=z)}{\mathbb{P}(Y=z)}$$

as above.

The equation (1) implies therefore the definition of cond. prob. in the discrete setting.

We will see next week that it also contains the example above with r.v. with joint prob. density. Indeed is the way to define a general notion of conditional probability.
