Stats 310A Session 4

November 1, 2019

1 Uniform integrability

The dominated convergence theorem states that $X_n \overset{a.s.}{\to} X_\infty$ and $|X_n| \leq Y$ for some integrable Y implies that $X_n \overset{L_1}{\to} X_\infty$ and $\mathbb{E}[X_n] \to \mathbb{E}[X_\infty]$. In this section, we explore *uniform integrability*, a useful concept that allows us to relax both conditions assumed in the dominated convergence theorem and still get L_1 convergence. We will follow Section 1.3.4 in Dembo's Notes.

Definition 1 (Uniform integrability)**.** *A collection of R.V.-s $\{X_\alpha, \alpha \in \mathcal{I}\}$ is called* uniformly integrable *(U.I.) if*

$$\lim_{n\to\infty} \sup_{\alpha\in\mathcal{I}} \mathbb{E}[|X_\alpha| \mathbf{1}\{|X_\alpha| > M\}] = 0.$$

Let us show that U.I. is indeed a relaxation of that $|X_\alpha| \leq Y$ for some integrable Y.

Lemma 1.1. *Let Y be integrable and suppose that $|X_\alpha| \leq Y$ for all α, then $\{X_\alpha\}$ is U.I.. In particular, any finite collection of integrable R.V.-s is U.I.. Further, if X_α is U.I. then $\sup_\alpha \mathbb{E}[|X_\alpha|] < \infty$.*

Proof That $\{X_\alpha\}$ is U.I. follows from that

$$\sup_\alpha \mathbb{E}[|X_\alpha| \mathbf{1}\{|X_\alpha| \geq M\}] \leq \mathbb{E}[|Y| \mathbf{1}\{|Y| \geq M\}] \to 0 \text{ as } M \to \infty.$$

If we have a finite collection $\{X_k\}_{k=1}^n$ that are integrable, then $Y = \sum_{k=1}^n |X_k|$ is integrable and dominates X_k. Suppose $\{X_\alpha\}$ is U.I., then for all M we have

$$\sup_\alpha \mathbb{E}[|X_\alpha|] \leq M + \sup_\alpha \mathbb{E}[|X_\alpha| \mathbf{1}\{|X_\alpha| > M\}].$$

The second term goes to zero as $M \to \infty$, so has to be finite for some M. This M yields a finite value on the RHS, so gives a finite upper bound on $\sup_\alpha \mathbb{E}[|X_\alpha|]$. □

To further understand U.I., consider the following example, which shows $\sup_\alpha \mathbb{E}[|X_\alpha|] < \infty$ does not necessarily give U.I..

Example 1: Let X_n be binary R.V.-s with $\mathbb{P}(X_n = 0) = 1 - 1/n$ and $\mathbb{P}(X_n = n) = 1/n$. Then $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] = 1$ for all n but X_n is not U.I.: for any M, take $n = \lceil M \rceil$, we have $\mathbb{E}[|X_n| \mathbf{1}\{|X_n| \geq M\}] = \mathbb{E}[|X_n|] = 1$. ◇

We are now ready to state the main convergence theorem.

Theorem 1 (Vitali's convergence theorem)**.** *Suppose $X_n \overset{p}{\to} X_\infty$, then the following are equivalent:*

(a) $\{X_n\}$ is U.I..

(b) $X_n \xrightarrow{L_1} X_\infty$.

(c) X_n *is integrable for all* $n \le \infty$ *and* $\mathbb{E}[|X_n|] \to \mathbb{E}[|X_\infty|]$.

Proof "(*a*) $\Longrightarrow$ (*b*)". We first deal with the case that $|X_n| \le M$ for some finite M. For all n and $\varepsilon > 0$, define

$$B_{n,\varepsilon} = \{\omega : |X_n(\omega) - X_\infty(\omega)| > \varepsilon\}.$$

As $X_n \xrightarrow{p} X_\infty$, we have $\mathbb{P}(B_{n,\varepsilon}) \to 0$ as $n \to \infty$ for all ε. In particular, we have $\mathbb{P}(|X_\infty| \ge M + \varepsilon) \le \mathbb{P}(B_{n,\varepsilon})$. Letting $n \to \infty$ gives $|X_\infty| \le M + \varepsilon$ almost surely, which after taking $\varepsilon \to 0$ gives that $|X_\infty| \le M$ almost surely. Hence, $|X_n - X_\infty| \le 2M$, which allows us to bound

$$\mathbb{E}[|X_n - X_\infty|] = \mathbb{E}[|X_n - X_\infty|\mathbf{1}\{B_{n,\varepsilon}\}] + \mathbb{E}[|X_n - X_\infty|\mathbf{1}\{B_{n,\varepsilon}^c\}] \le 2M\mathbb{P}(B_{n,\varepsilon}) + \varepsilon.$$

Taking $n \to \infty$, we conclude that $\limsup_{n\to\infty} \mathbb{E}[|X_n - X_\infty|] \le \varepsilon$, so as ε is arbitrary we get $\mathbb{E}[|X_n - X_\infty|] \to 0$.

We now show the general case where $\{X_n\}$ is U.I. by applying a truncation argument. Define the truncation function

$$\varphi_M(x) = x\mathbf{1}\{|x| \le M\}.$$

As $X_n \xrightarrow{p} X_\infty$ and φ_M is continuous, we have $\varphi_M(X_n) \xrightarrow{p} \varphi_M(X_\infty)$. The R.V.-s $\varphi_M(X_n)$ are bounded in $[-M, M]$, so by the bounded case, we get $\mathbb{E}[|\varphi_M(X_n) - \varphi_M(X_\infty)|] \to 0$.

As $\{X_n\}$ is U.I., we have $\sup_n \mathbb{E}[|X_n|] = c < \infty$ by Lemma 1.1, which gives that

$$c = \sup_n \mathbb{E}[|X_n|] \ge \sup_n \mathbb{E}[|\varphi_M(X_n)|] \ge \lim_{n\to\infty} \mathbb{E}[|\varphi_M(X_n)|] = \mathbb{E}[\varphi_M(X_\infty)].$$

The R.V.-s $|\varphi_M(X_\infty)|$ is an increasing sequence as $M \uparrow \infty$ and converges to $|X_\infty|$. By the monotone convergence theorem, we have

$$\mathbb{E}[|X_\infty|] = \lim_{M\to\infty} \mathbb{E}[|\varphi_M(X_\infty)|] \le c,$$

so X_∞ is integrable.

As $\{X_n\}$ is U.I. and X_∞ is integrable, for any $\varepsilon > 0$, there exists some M such that

$$\sup_n \mathbb{E}[|X_n|\mathbf{1}\{|X_n| \ge M\}] \le \varepsilon,$$

and $\mathbb{E}[|X_\infty|\mathbf{1}\{|X_\infty| \ge M\}] \le \varepsilon$. By the triangle inequality, we have

$$\begin{aligned}\mathbb{E}[|X_n - X_\infty|] &\le \mathbb{E}[|X_n - \varphi_M(X_n)|] + \mathbb{E}[|\varphi_M(X_n) - \varphi_m(X_\infty)|] + \mathbb{E}[|\varphi_M(X_\infty) - X_\infty|] \\ &\le 2\varepsilon + \mathbb{E}[|\varphi_M(X_n) - \varphi_M(X_\infty)|].\end{aligned}$$

Letting $n \to \infty$, we get $\limsup_{n\to\infty} \mathbb{E}[|X_n - X_\infty|] \le 2\varepsilon$. Taking $\varepsilon \downarrow 0$ gives that $\mathbb{E}[|X_n - X_\infty|] \to 0$, the desired result.

That (*b*) $\Longrightarrow$ (*c*) is immediate, and we will skip the proof of (*c*) $\Longrightarrow$ (*a*). $\square$

2 Kolmogorov's extension theorem

We state and prove the Kolmogorov's extension theorem when the index set is $T = \{1, 2, 3, \ldots\} = \mathbb{N}$.

Theorem 2 (Theorem 1.4.22, Dembo's Notes)**.** *Suppose we are give probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ that are consistent, that is,*

$$\mu_{n+1}(B_1 \times \cdots \times B_n \times \mathbb{R}) = \mu_n(B_1 \times \cdots \times B_n) \quad \forall B_i \in \mathcal{B},\ i = 1, \ldots, n < \infty. \tag{1}$$

Then, there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_c)$ such that

$$\mathbb{P}(\{\omega : \omega_i \in B_i, i = 1, \ldots, n\}) = \mu_n(B_1 \times \cdots \times B_n) \quad \forall B_i \in \mathcal{B},\ i = 1, \ldots, n < \infty.$$

Remark Kolmogorov's extension theorem builds the foundation on which stochastic processes are defined: namely, for any index set T, to define the distribution of a stochastic process X_T, it suffices to give a consistent collection of joint distributions of $(X_{t_1}, \ldots, X_{t_n})$ on finitely many coordinates. The measure of X_T on $(\mathbb{R}^T, \mathcal{B}_c)$, then, by the extension theorem, is guaranteed to exist and is unique.

The theorem is trivial when $T = \{1, \ldots, n\}$ is finite: just take $\mathbb{P} = \mu_n$. $T = \mathbb{N}$ is the first non-trivial case of the theorem. This case can give us, for example, the probability measure of countably many i.i.d. R.V.-s $(X_1, X_2, \ldots)$.

Proof of Theorem 2 The proof mainly follows that of [1, Chapter 36]. Let $\mathbb{R}_0^{\mathbb{N}}$ be the collection of cylindral sets of the form

$$A = \Big\{x \in \mathbb{R}^{\mathbb{N}} : (x_1, \ldots, x_n) \in H\Big\}, \tag{2}$$

where $n \in \mathbb{N}$ and $H \in \mathcal{B}_{\mathbb{R}^n}$. That is, we consider sets that require the first n coordinates lie in some Borel set $H \subset \mathbb{R}^n$. By definition of the cylindral σ-algebra, we have $\mathcal{B}_c = \sigma(\mathbb{R}_0^{\mathbb{N}})$. On this collection, define the set function

$$\mathbb{P}(A) = \mu_n(H).$$

We are going to use Caratheodory's extension theorem to extend $\mathbb{P}$ to $\mathcal{B}_c$, which we divide into the following steps.

$\mathbb{P}$ is well-defined To show this, we need to verify that if a cylindral set A has two representations of the form (2) then they give coinciding values of $\mathbb{P}(A)$. Consider

$$A = \{x : (x_1, \ldots, x_{n_1}) \in H_1\} = \{x : (x_1, \ldots, x_{n_2}) \in H_2\}$$

for some $n_1 \geq n_2$, then it is easy to see that $H_1 = H_2 \times \mathbb{R}^{n_1 - n_2}$. (Check this!) It remains to show that

$$\mu_{n_1}(H_1) = \mu_{n_1}(H_2 \times \mathbb{R}^{n_1 - n_2}) = \mu_{n_2}(H_2). \tag{3}$$

Repeating the consistency condition (1) gives that $\mu_{n_1}(B_1 \times \cdots \times B_{n_2} \times \mathbb{R}^{n_1 - n_2}) = \mu_{n_2}(B_1 \times \cdots \times B_{n_2})$, and a standard extension argument shows that $\mu_{n_1}(\cdot \times \mathbb{R}^{n_1 - n_2}) = \mu_{n_2}(\cdot)$, verifying (3).

$\mathbb{R}_0^{\mathbb{N}}$ is an algebra; $\mathbb{P}$ finitely additive on $\mathbb{R}_0^{\mathbb{N}}$ Clearly $\emptyset \in \mathbb{R}_0^{\mathbb{N}}$. For any cylindral set A, we have $A^c = \left\{x \in \mathbb{R}^{\mathbb{N}} : (x_1, \ldots, x_n) \in H^c\right\}$, so $A^c \in \mathbb{R}_0^{\mathbb{N}}$. Let A, B be two cylindral sets:

$$A = \{x : (x_1, \ldots, x_{n_1}) \in H_1\}, \quad B = \{x : (x_1, \ldots, x_{n_2}) \in H_2\}.$$

Without loss of generality, let $n_1 \geq n_2$. We then have

$$A \cup B = \left\{x : (x_1, \ldots, x_{n_1}) \in H_1 \cup (H_2 \times \mathbb{R}^{n_1 - n_2})\right\} \in \mathbb{R}_0^{\mathbb{N}}. \tag{4}$$

This shows that $\mathbb{R}_0^{\mathbb{N}}$ is an algebra. If A and B are disjoint, then $H_2 \times \mathbb{R}^{n_1-n_2} \cap H_1 = \emptyset$, giving that

$$\mathbb{P}(A \cup B) = \mu_{n_1}(H_1 \cup (H_2 \times \mathbb{R}^{n_1-n_2})) = \mu_{n_1}(H_1) + \mu_1(H_2 \times \mathbb{R}^{n_1-n_2}) = \mathbb{P}(A) + \mathbb{P}(B),$$

so $\mathbb{P}$ is finitely additive.

$\mathbb{P}$ is a probability measure on $\mathbb{R}_0^{\mathbb{N}}$ Clearly $\mathbb{P} \geq 0$ and $\mathbb{P}(\emptyset) = 0$. Let A be a cylindral set, then

$$\mathbb{P}(A^c) = \mu_n(H^c) = 1 - \mu_n(H) = 1 - \mathbb{P}(A).$$

It remains to show countable additivity. As it is finitely additive, it suffices to show that $A_k \in \mathbb{R}_0^{\mathbb{N}}$ with $A_k \downarrow \emptyset$ implies $\mathbb{P}(A_k) \to 0$. (See the Remark in Dembo notes, page 14). As we can always make the defining index non-decreasing, we can let

$$A_k = \{x : (x_1, \ldots, x_{n_k}) \in H_k\}$$

where $n_k \in \mathbb{N}$ is increasing and $H_k \subset \mathbb{R}^{n_k}$.

Suppose $\mathbb{P}(A_k) \not\to 0$, then $\mathbb{P}(A_k) \geq \varepsilon$ holds for all k, for some $\varepsilon > 0$. This means $\mu_{n_k}(H_k) \geq \varepsilon$. Applying [1, Theorem 12.3], there exists compact sets $K_k \subseteq H_k$ such that $\mu_{n_k}(H_k \setminus K_k) \leq \varepsilon/2^{k+1}$. Define

$$B_k = \{x : (x_1, \ldots, x_{n_k}) \in K_k\},$$

then $\mathbb{P}(A_k \setminus B_k) \leq \varepsilon/2^{k+1}$. Define $C_k = \bigcap_{j=1}^k B_j$, then we have $C_k \subset B_k \subset A_k$ and $\mathbb{P}(A_k \setminus C_k) \leq \varepsilon/2$, so $\mathbb{P}(C_k) \geq \varepsilon/2$, and thus C_k is non-empty.

Now, for all k, choose a point $x^{(k)} \in C_k$. As C_k is the intersection of $\{B_j\}_{j \leq k}$, we have $(x_1^{(k)}, \ldots, x_{n_j}^{(k)}) \in K_j$ for all $j \leq k$. In other words, the first n_j indices of $\left\{x^{(k)}\right\}_{k \geq j}$ lie in the compact set K_j. Hence, there exists a subsequence k_i such that $(x_1^{(k_i)}, \ldots, x_{n_j}^{(k_i)})$ converges. By the diagonal method, we can find a subsequence k_i such that $(x_1^{(k_i)}, \ldots, x_{n_j}^{(k_i)})$ converges for all j. Let x be the point in $\mathbb{R}^{\mathbb{N}}$ such that $(x_1, \ldots, x_{n_j})$ is the limit of the above sequence (as the limits are consistent, x exists). The closedness of K_j implies that $(x_1, \ldots, x_{n_j}) \in K_j$, so $x \in A_j$. Thus we have found a point $x \in \bigcap_{j=1}^{\infty} A_j$, contradictory to that $A_j \downarrow \emptyset$. Hence our assumption is wrong so we must have $\mathbb{P}(A_j) \to 0$. $\square$

References

[1] P. Billingsley. *Probability and measure.* John Wiley & Sons, 2008.