STATS300B – Lecture 3

Julia Palacios
Scribed by Michael Howes

01/11/22

Contents

1 Portmanteau theorem

Last time we stated the Portmanteau theorem.

Theorem 1. *Let X_n and X be a random vectors. The following are equivalent.*

1. $X_n \stackrel{d}{\to} X$.
2. $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all bounded and continuous f.*
3. $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all Lipschitz f with $f(x) \in [0,1]$ for all x.*
4. $\liminf_n \mathbb{E}[f(X_n)] \geq \mathbb{E}[f(X)]$ *for all continuous non-negative f.*
5. $\liminf_n \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$ *for all open sets O.*
6. $\limsup_n \mathbb{P}(X_n \in C) \leq \mathbb{P}(X \in C)$ *for all closed sets C.*
7. $\lim_n \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$ *for all measurable sets B such that $\mathbb{P}(X \in \delta B) = 0$ where δB denotes the boundary of B.*

We will not prove the full theorem, but we will prove some parts to give the flavor of the arguments. Today we will prove that if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all bounded and continuous f, then $\limsup_n \mathbb{P}(X_n \in C) \leq \mathbb{P}(X \in C)$.

Proof. Assume $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all bounded and continuous $f : \mathbb{R}^d \to \mathbb{R}$ and let $C \subseteq \mathbb{R}^d$ be a closed set. Consider the function $h_C : \mathbb{R}^d \to [0, \infty)$ given by

$$h_C(x) = \inf\{\|x - y\| : y \in C\}.$$

Since C is closed, we have $h_C(x) = 0$ if and only if $x \in C$. The function h_C is continuous. For each $J \in \mathbb{N}$ define $\phi_J : \mathbb{R} \to \mathbb{R}$ by

$$\phi_J(t) = \begin{cases} 1 & \text{if } t \leq 0, \\ 1 - Jt & \text{if } 0 < t < \frac{1}{J}, \\ 0 & \text{if } t \geq \frac{1}{J}. \end{cases}$$

Also define $f_J(x) = \phi_J(h_C(x))$. The functions f_J are continuous and bounded. Furthermore, $f_J(x) \to \mathbf{1}_C(x)$ and $f_J(x) \geq \mathbf{1}_C(x)$ for all $x \in \mathbb{R}^d$. Thus, for all J,

$$\begin{aligned} \mathbb{P}(X_n \in C) &= \mathbb{E}[\mathbf{1}_C(X_n)] \\ &\leq \mathbb{E}[f_J(X_n)]. \end{aligned}$$

By taking n to infinity, we have $\limsup_n \mathbb{P}(X_n \in C) \leq \mathbb{E}[f_J(X)]$. But $|f_J(x)| \leq 1$ and $f_J(X)$ converges to $\mathbf{1}_C(X)$. By the dominated convergence theorem we therefore have

$$\lim_{J\to\infty} \mathbb{E}[f_J(X)] = \mathbb{E}[\mathbf{1}_C(X)] = \mathbb{P}(X \in C).$$

Therefore,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in C) \geq \mathbb{P}(X \in C) \qquad \square$$

Definition 1. A collection of functions $\mathcal{F}$ is a *convergence determining class* if for all random vectors $(X_n)_{n\geq 1}$ and X, $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ if and only if $X_n \xrightarrow{d} X$.

The Portmanteau Theorem show that all bounded and continuous functions is a convergence determining class. As is the class of Lipschitz functions taking values in $[0, 1]$. Another important class of convergence determining functions are the functions

$$f_t(x) = e^{it\cdot x},$$

which is a class indexed by t. The function

$$\phi_X(t) = \mathbb{E}[f_t(X)] = \mathbb{E}[e^{it\cdot X}],$$

is called the characteristic function of X. Since $\{f_t\}_{t\in\mathbb{R}^d}$ is a convergence determining class, we know that $X_n \xrightarrow{d} X$ if and only if $\phi_{X_n}(t) \to \phi_X(t)$ for all t.

Note that the assumption that f is bounded is important. A sequence X_n may converge in distribution to X, but this does not imply that $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for continuous unbounded f. Indeed, we may not have $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

2 Tightness

Definition 2. A collection of random vectors $\{X_a\}_{a\in\mathcal{A}}$ is *uniformly tight* if for all $\varepsilon > 0$, there exists $M < \infty$ such that

$$\sup_{a\in\mathcal{A}} \mathbb{P}(\|X_a\| > M) \leq \varepsilon.$$

A uniformly tight collection of random vectors is sometimes said to be *bounded in probability.* This is because if $\{X_a\}_{a\in\mathcal{A}}$ is uniformly tight, then with probability at least $1 - \varepsilon$, $\|X_a\| \leq M$ for every a.

We can also define uniform tightness for probability measures instead of random variables.

Definition 3. A collection of probability measures $\{\mathbb{P}_a\}_{a\in\mathcal{A}}$ on $\mathbb{R}^d$ is *uniformly tight* if for all $\varepsilon > 0$, there exists a compact set C such that

$$\sup_{a\in\mathcal{A}} \mathbb{P}_a(C) \geq 1 - \varepsilon.$$

Remark 1. The following are examples of uniformly tight collections.

1. A single random vector X is uniformly tight since
$$\lim_{n\to\infty} \mathbb{P}(\|X\| \le n) = \mathbb{P}(\|X\| < \infty) = 1.$$

2. If $\{X_a\}_{a\in\mathcal{A}_1}, \{X_a\}_{a\in\mathcal{A}_2}, \dots \{X_a\}_{a\in\mathcal{A}_m}$ are all uniformly tight, then $\{X_a\}_{a\in\bigcup_{i=1}^m \mathcal{A}_i}$ is also uniformly tight.

3. If $X_n \xrightarrow{d} X$, then $\{X_n\}_{n\ge1}$ is uniformly tight.

The converse of the last remark is almost true. A uniformly tight collection of random vectors need not converge in distribution, but there must be a subsequence which does.

Theorem 2. *If $\{X_n\}_{n\ge1}$ is uniformly tight, then there exists a random vector X and a subsequence n_j such that $X_{n_j} \xrightarrow{d} X$.*

Note that the original sequence $\{X_n\}$ need not converge to anything as the following example shows.
$$X_n = \begin{cases} 1+\frac{1}{n} & \text{if } n \text{ is odd,} \\ 2+\frac{1}{n} & \text{if } n \text{ is even.} \end{cases}$$

3 Convergence in L^p

Recall that $X_n \xrightarrow{L^p} X$ if $\mathbb{E}[\|X_n - X\|_p^p] \to 0$. The following links convergence in L^p to convergence in distribution and convergence in probability.

Definition 4. A sequence $\{X_n\}_{n\ge1}$ is *uniformly integrable* if
$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{E}[|X_n| \mathbf{1}_{|X_n|\ge\lambda}] = 0.$$

If $X_n \xrightarrow{d} X$, then for every $r > 0$, $\mathbb{E}[\|X_n\|_r^r] \to \mathbb{E}[\|X\|_r^r]$ if and only if $\{\|X_n\|_r^r\}_{n\ge1}$ is uniformly integrable.

Theorem 3 (Vitalli)**.** *Suppose $X_n \in L^r$ for some $r \in (0,\infty)$ and that $X_n \xrightarrow{p} X$. Then the following are equivalent,*

1. *$\{\|X_n\|_r^r\}$ are uniformly integrable.*
2. *$X_n \xrightarrow{L^r} X$*
3. *$\limsup_n \mathbb{E}\,\|X_n\|_r^r \le \mathbb{E}\,\|X\|_r^r$.*

4 Almost sure convergence

Definition 5. A sequence of random variables $\{X_n\}$ *converge almost surely* to X if
$$\mathbb{P}\left(\lim_{n\to\infty} X_n \ne X\right) = 0.$$

We denote almost sure convergence by $X_n \xrightarrow{a.s.} X$.

The following are equivalent

1. $X_n \overset{a.s.}{\to} X$.
2. For all $\varepsilon > 0$,
$$\mathbb{P}(\|X_n - X\| > \varepsilon, \text{ infinitely often}) = \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} \|X_m - X\| > \varepsilon\right) = 0.$$
3. For all $\varepsilon > 0$,
$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{m=n}^{\infty} \|X_m - X\| > \varepsilon\right) = 0.$$
4. For all $\varepsilon > 0$,
$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{m\geq n} \|X_m - X\| \geq \varepsilon\right) = 0.$$

The following theorem is called the strong law of large numbers.

Theorem 4. *Suppose $X_1, \ldots$ are i.i.d. with $\mathbb{E}|X_i| < \infty$ and $\mathbb{E}[X_i] = \mu$. Then $\bar{X}_n \overset{a.s.}{\to} \mu$.*

The Borel–Cantelli lemmas are the main tools for proving almost sure convergence.

Proposition 1. *1. Let $\{A_n\}_{n\geq 1}$ be any sequence of events. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then*
$$\mathbb{P}(A_n \textit{ infinitely often}) = \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{m=1}^{\infty} A_m\right) = 0.$$

2. If $\{A_n\}_{n\geq 1}$ is a sequence of independent events and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then
$$\mathbb{P}(A_n \textit{ infinitely often}) = 1.$$

Proof. We will only prove 1. Note that, for all $n \in \mathbb{N}$,
$$\mathbb{P}(A_n \text{ infinitely often}) \leq \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m).$$
Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, the sum $\sum_{m=n}^{\infty} \mathbb{P}(A_m)$ goes to zero as n goes to infinity. Thus,
$$\mathbb{P}(A_n \text{ infinitely often}) = 0.$$
□

5 Standard implications

We have the following implications
$$X_n \overset{a.s.}{\to} X \Longrightarrow X_n \overset{p}{\to} X \Longrightarrow X_n \overset{d}{\to} X,$$
and for any $p > 0$,
$$X_n \overset{L^p}{\to} X \Longrightarrow X_n \overset{p}{\to} X \Longrightarrow X_n \overset{d}{\to} X.$$
The converse implications do not hold in general, but partial converses do hold. For instance if b is a constant, then
$$X_n \overset{d}{\to} b \Longrightarrow X_n \overset{p}{\to} b.$$
Also, if $X_n \overset{p}{\to} X$, then there exists a subsequence X_{n_j} such that $X_{n_j} \overset{a.s.}{\to} X$. Likewise, if $X_n \overset{L^p}{\to} X$, then there exists a subsequence X_{n_j} such that $X_{n_j} \overset{a.s.}{\to} X$. Also, if $X_n \overset{a.s.}{\to} X$ and $\{\|X_n\|_p^p\}$ are uniformly

integrable, then $X_n \xrightarrow{L^p} X$. But in general, almost sure convergence does not imply convergence in L^p and convergence in L^p does not imply convergence almost surely.

We have already proven some of these implications are others are given as homework. We will prove a few more now. Firstly we will show

$$X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{p} X.$$

Proof. Suppose $X_n \xrightarrow{a.s.} X$ and let $\varepsilon > 0$. Then

$$\mathbb{P}(\|X_n - X\| > \varepsilon) \leq \mathbb{P}\left(\bigcup_{m=n}^{\infty} \|X_m - X\| > \varepsilon\right).$$

By almost sure convergence, the term on the right goes to 0. Thus, $\mathbb{P}(\|X_n - X\| > \varepsilon) \to 0$ and so $X_n \xrightarrow{p} X$. $\square$

We will also prove

$$X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X.$$

Proof. Suppose $X_n \xrightarrow{p} X$ and let t be a continuity point of F, the cumulative distribution function of X. Let F_n be the cumulative distribution function of X_n. Fixing $\varepsilon > 0$, we have

$$\begin{aligned} F_n(t) &= \mathbb{P}(X_n \leq t) \\ &= \mathbb{P}(X_n \leq t, X \leq t + \varepsilon) + \mathbb{P}(X_n \leq t, X > t + \varepsilon) \\ &\leq \mathbb{P}(X \leq t + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon) \\ &= F(t + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon) \end{aligned}$$

Since $X_n \xrightarrow{p} X$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$. Thus,

$$\limsup_{n\to\infty} F_n(t) \leq F(t + \varepsilon).$$

Similarly,

$$\begin{aligned} F(t - \varepsilon) &= \mathbb{P}(X \leq t - \varepsilon) \\ &= \mathbb{P}(X \leq t - \varepsilon, X_n \leq t) + \mathbb{P}(X \leq t - \varepsilon, X_n \geq t) \\ &\leq \mathbb{P}(X_n \leq t) + \mathbb{P}(|X_n - X| \geq \varepsilon) \\ &= F_n(t) + \mathbb{P}(|X_n - X| \geq \varepsilon). \end{aligned}$$

Thus,

$$F_n(t) \geq F(t - \varepsilon) - \mathbb{P}(|X_n - X| \geq \varepsilon).$$

Which implies,

$$\liminf_{n\to\infty} F_n(t) \geq F(t - \varepsilon).$$

Since F is continuous at t, both $F(t-\varepsilon)$ and $F(t+\varepsilon)$ can be made arbitrarily close to $F(t)$ and hence

$$\limsup_{n\to\infty} F_n(t) \leq F(t) \leq \liminf_{n\to\infty} F_n(t).$$

Thus $\lim\limits_{n\to\infty} F_n(t) = F(t)$. $\square$