STATS300B – Lecture 4

Julia Palacios
Scribed by Michael Howes

01/13/22

Contents

1 Relationships between the modes of convergence

We ended last lecture with the statement of the following implications,

$$X_n \stackrel{a.s.}{\to} X \implies X_n \stackrel{p}{\to} X \implies X_n \stackrel{d}{\to} X,$$

and for any $r > 0$,

$$X_n \stackrel{L^r}{\to} X \implies X_n \stackrel{p}{\to} X \implies X_n \stackrel{d}{\to} X.$$

None of the converse are true in generality, but today we will see some partial converses. Starting with the following,

Proposition 1. *Suppose that $X_n \stackrel{p}{\to} X$, then there exists a subsequence X_{n_k} such that $X_{n_k} \stackrel{a.s.}{\to} X$ as $k \to \infty$.*

Proof. Suppose that $X_n \stackrel{p}{\to} X$. Then for every $k \in \mathbb{N}$, there exists n_k such that,

$$\mathbb{P}(\|X_{n_k} - X\| \geq 1/k) \leq 2^{-k}.$$

We may choose the integers n_k so that they are strictly increasing in k. We will now show that $X_{n_k} \stackrel{a.s.}{\to} X$. Let B be the set on which $X_{n_k} \to X$. Note that

$$A_m = \bigcap_{k=m}^{\infty} \{\|X_{n_k} - X\| < 1/k\} \subseteq B.$$

Furthermore,

$$\mathbb{P}(A_m^C) \leq \sum_{k=m}^{\infty} \mathbb{P}(\|X_{n_k} - X\| \geq 1/k) \leq \sum_{k=m}^{\infty} \frac{1}{2^k}.$$

Thus, $\mathbb{P}(A_m^C) \to 0$ as $m \to \infty$ and hence $\mathbb{P}(A_m) \to 1$ as $m \to \infty$. Since $\mathbb{P}(B) \geq \mathbb{P}(A_m)$ for every m we have $\mathbb{P}(B) = 1$. $\square$

2 Scales of magnitude

Recall the following definitions for describing the asymptotic relationship between sequences of numbers.

Definition 1. Let $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$ be sequences of constants. Then,

1. $a_n = o(b_n)$ means that $\frac{a_n}{b_n} \to 0$ as $n \to \infty$.
2. $a_n = O(b_n)$ means that $\limsup_{n\to\infty} \left|\frac{a_n}{b_n}\right| < \infty$.
3. $a_n \sim b_n$ means that $\frac{a_n}{b_n} \to 1$ as $n \to \infty$.

These definitions all have probabilistic analogs for sequences of random variables.

Definition 2. Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables and let $\{b_n\}_{n\geq 0}$ be a sequence of constants. Then,

1. $X_n = o_p(b_n)$ means that $\frac{X_n}{a_n} \xrightarrow{p} 0$ as $n \to \infty$.
2. $X_n = O_p(1)$ means that
$$\lim_{K\to\infty} \sup_n \mathbb{P}(|X_n| \geq K) = 0.$$
3. $X_n = O_p(b_n)$ means that $\frac{X_n}{b_n} = O_p(1)$.

The following arithmetic rules are useful and simple to prove.

Lemma 1. *We have*

$$\begin{aligned} o_p(1) + o_p(1) &= o_p(1) \\ O_p(1) + O_p(1) &= O_p(1) \\ O_p(a_n)O_P(b_n) &= O_P(a_nb_n) \\ O_p(a_n)o_p(b_n) &= o_p(a_nb_n). \end{aligned}$$

3 Inequalities for the L^r space

We will now state and prove some important inequalities and facts about random variables in L^r.

Proposition 2. *If $\mathbb{E}|X|^r < \infty$, then $\mathbb{E}|X|^{r'} < \infty$ for all $r' \leq r$.*

Proof. If $r' \leq r$, then $|X|^{r'} \leq 1 + |X|^r$. Thus, if $\mathbb{E}|X|^r < \infty$, then

$$\mathbb{E}|X|^{r'} \leq 1 + \mathbb{E}|X|^r < \infty$$ □

Proposition 3. *For any random variable X, $\mathrm{Var}(X) < \infty$ if and only if $\mathbb{E}[X^2] < \infty$.*

Proof. If $\mathbb{E}[X^2] < \infty$, then $\mathbb{E}[X] \leq \mathbb{E}[|X|] < \infty$. And thus

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 < \infty.$$

If $\mathrm{Var}(X) < \infty$, then $\mathbb{E}[(X - \mathbb{E}[X])^2] < \infty$ in particular $\mathbb{E}[X] \in \mathbb{R}$. Thus, $\mathbb{E}[X^2] = \mathbb{E}[X]^2 + \mathrm{Var}(X) < \infty$. □

Proposition 4. *For every $r > 0$, $\mathbb{E}|X + Y|^r \leq c_r\mathbb{E}|X|^r + c_r\mathbb{E}|Y|^r$ where $c_r = 1$ if $0 < r \leq 1$ and $c_r = 2^{r-1}$ for $r \geq 1$.*

Proof. First suppose that $r > 1$. The function $f(x) = |x|^r$ is convex and thus,

$$\left|\frac{X+Y}{2}\right|^r \leq \frac{1}{2}|X|^2 + \frac{1}{2}|Y|^r.$$

Thus, $|X+Y|^r \leq 2^{r-1}|X|^r + 2^{r-1}|Y|^r$.

Now suppose $0 < r \leq 1$. If $|X+Y|^r \leq |X|^r$, then we are done. If $|X+Y|^r > |X|^r$, then

$$\begin{aligned}
|X+Y|^r - |X|^r &= \int_{|X|}^{|X+Y|} rt^{r-1}dt \\
&\leq \int_{|X|}^{|X|+|Y|} rt^{r-1}dt \\
&= \int_0^{|Y|} r(t+|Y|)^{r-1}dt \\
&\leq \int_0^{|y|} rt^{r-1}dt \\
&= |Y|^r.
\end{aligned}$$

Thus, $|X+Y|^r \leq |X|^r + |Y|^r$. $\square$

We next prove Hölder's inequality.

Proposition 5. *Let $r, s \geq 1$ be such that $\frac{1}{r} + \frac{1}{s} = 1$, then*

$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^r)^{1/r} \left(\mathbb{E}^{1/s}|Y|^s\right)^{1/s}.$$

Proof. If $\mathbb{E}|X|^r = 0$ or $\mathbb{E}|Y|^s = 0$, then $X = 0$ almost surely or $Y = 0$ almost surely. Hence, $XY = 0$ almost surely and thus $\mathbb{E}|XY| = 0$. Thus, we may assume $\mathbb{E}|X|^r, \mathbb{E}|Y|^s > 0$. If $\mathbb{E}|X|^r = \infty$ or $\mathbb{E}|Y|^s = \infty$, then we are done. Thus, we will assume that $\mathbb{E}|X|^r, \mathbb{E}|Y|^s \in (0, \infty)$. We will first prove Young's inequality which states for all $a, b \geq 0$,

$$ab \leq \frac{a^r}{r} + \frac{b^s}{s}.$$

Note that if $a = 0$ or $b = 0$, then Young's inequality is an equality. Thus assume that $a, b > 0$. We know that the function $x \mapsto e^x$ is convex. Since $\frac{1}{r} + \frac{1}{s} = 1$, we thus have

$$\begin{aligned}
ab &= e^{\log(ab)} \\
&= e^{\log(a)+\log(b)} \\
&= e^{\frac{1}{r}\log(a^r)+\frac{1}{s}\log(b^r)} \\
&\leq \frac{1}{r}e^{\log(a^r)} + \frac{1}{s}e^{\log(b^s)} \\
&= \frac{a^r}{r} + \frac{b^s}{s},
\end{aligned}$$

as claimed. We will now apply Young's inequality point-wise to prove Hölder's inequality,

$$\begin{aligned}
\frac{\mathbb{E}[|XY|]}{\left(\mathbb{E}|X|^r\right)^{1/r}\left(\mathbb{E}^{1/s}|Y|^s\right)^{1/s}} &= \int_\Omega \frac{|X||Y|}{\left(\mathbb{E}|X|^r\right)^{1/r}\left(\mathbb{E}^{1/s}|Y|^s\right)^{1/s}} d\mathbb{P} \\
&\leq \int_\Omega \frac{1}{r}\frac{|X|^r}{\mathbb{E}|X|^r} + \frac{1}{s}\frac{|Y|^s}{\mathbb{E}|Y|^s} d\mathbb{P} \\
&= \frac{1}{r\mathbb{E}|X|^r}\int_\Omega |X|^r d\mathbb{P} + \frac{1}{s\mathbb{E}|Y|^s}\int_\Omega |Y|^s d\mathbb{P} \\
&= \frac{1}{r} + \frac{1}{s} \\
&= 1.
\end{aligned}$$

Thus ,$\mathbb{E}|XY| \leq \left(\mathbb{E}|X|^r\right)^{1/r}\left(\mathbb{E}^{1/s}|Y|^s\right)^{1/s}$. □

4 Convergence in distribution

The following theorem shows that we can study convergence in distribution of random vectors by projecting onto one dimensional subspaces.

Theorem 1 (Cramér–Wold device)**.** *Let X_n and X be random vectors in $\mathbb{R}^d$, then $X_n \xrightarrow{d} X$ if and only if $a^T X_n \xrightarrow{d} a^T X$ for all constants $a \in \mathbb{R}^d$.*

We also have a version of the continuous mapping theorem for convergence in distribution and almost sure convergence. The version for convergence in probability was stated in the first lecture.

Theorem 2 (Continuous mapping theorem)**.** *Let g be a continuous function on a set B such that $\mathbb{P}(X \in B) = 1$. Then*

1. *If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{p} g(X)$.*
2. *If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$.*
3. *If $X_n \xrightarrow{a.s.} X$, then $g(X_n) \xrightarrow{d} g(X)$.*

We have already proved 1. For now, we will only prove 3. Once we have Skorokhod's theorem we will see that 3 implies 2.

Proof. Let A be the set on which $X_n \to X$. Since g is continuous on B, we have $g(X_n) \to g(X)$ on $A \cap B$. Since A and B both have probability 1, $\mathbb{P}(A \cap B) = 1$ and thus $\mathbb{P}(g(X_n) \to g(X)) = 1$. □

Definition 3. Given a cumulative distribution function F on $\mathbb{R}$, define $F^{-1} : (0, 1) \to \mathbb{R}$ to be the function

$$F^{-1}(t) = \inf\{x : F(x) \geq t\}.$$

The function F^{-1} is called the *quantile function* of F.

Proposition 6. *The function F^{-1} is non-decreasing and left-continuous. And for all $t \in (0, 1)$ and $x \in \mathbb{R}$,*

$$F^{-1}(t) \leq x \iff t \leq F(x).$$

Proof. The set $\{x : F(x) \geq t\}$ are non-increasing with t and thus if $t \leq t'$, then

$$F^{-1}(t) = \inf\{x : F(x) \geq t\} \leq \inf\{x : F(x) \geq t'\} = F^{-1}(t').$$

Showing that F^{-1} is non-increasing. Continuity can be proved by considered first the points t such that F is continuous at $F^{-1}(t)$ and then the points t where F is discontinuous at $F^{-1}(t)$. A picture helps.

Finally, if $F^{-1}(t) \leq x$, then $\inf\{z : F(z) \geq t\} \leq x$. Thus, for any $\varepsilon > 0$, there exists $z < x + \varepsilon$ such that $F(z) \leq t$. Since F is right-continuous, this implies that

$$F(x) = \lim_{z \searrow x} F(z) \leq t.$$

Conversely, if $t \leq F(x)$, then $x \in \{z : F(z) \geq t\}$ and so $F^{-1}(t) \leq x$. □

The function F^{-1} also have the following properties.

Proposition 7. *Let X be random variable with CDF F. Then for all $t \in (0,1)$,*

$$\mathbb{P}(F(X) \leq t) \leq t,$$

and we have equality if and only if t is in the range of F. In particular if F is continuous, then the above holds for all $t \in (0,1)$ and thus $F(X) \sim U(0,1)$. We can also write the above inequality as for all $t \in (0,1)$

$$F(F^{-1}(t)) \geq t.$$

We also have $F^{-1}(F(x)) \leq x$ for all $x \in \mathbb{R}$ with strict inequality if and only if $F(x - \varepsilon) = F(x)$ for some $\varepsilon > 0$. Thus, $\mathbb{P}(F^{-1}(F(X)) \neq X) = 0$.

We also have

Proposition 8. *Let $U \sim U(0,1)$ and let F be some CDF function. Let $X = F^{-1}(U)$. Then $\{X \leq x\} = \{U \leq F(x)\}$ and so $X \sim F$.*

We will now state and prove *Skorokhod's representation theorem.*

Theorem 3. *Suppose that $X_n \stackrel{d}{\to} X$, then there exist random variables X_n^* and X^* such that $X_n^* \stackrel{a.s.}{\to} X^*$, $X_n^* \stackrel{\text{dist}}{=} X_n$ and $X^* \stackrel{\text{dist}}{=} X$.*

Proof. Let F_n be the CDF of X_n and let F be the CDF of X. Define $X_n^* = F_n^{-1}(U)$ and $X^* = F^{-1}(U)$. We will show that for all but a countable number of $t \in (0,1)$ we have $F_n^{-1}(t) \to F^{-1}(t)$. This will imply that $X_n^* \to X^*$ with probability 1.

Since F^{-1} is increasing, F^{-1} has at most countably many discontinuities. Let $t \in (0,1)$ be a point such that F^{-1} is continuous at t and let $\varepsilon > 0$. We can find a value x such that F is continuous at x and

$$F^{-1}(t) - \varepsilon < x < F^{-1}(t).$$

It follows that $F(x) < t$. Since F is continuous at x we know that $F_n(x) \to F(x)$. Thus, for large enough n, $F_n(x) < t$ which implies $x \leq F_n^{-1}(t)$ and so

$$\liminf_n F_n^{-1}(t) \geq x \geq F^{-1}(t) - \varepsilon.$$

Which implies $\liminf_n F_n^{-1}(t) \geq F^{-1}(t)$. Now consider $s > t$ and choose y such that $F^{-1}(s) < y < F^{-1}(s) + \varepsilon$ and F is continuous at y. Thus, $t < s \leq F(y)$. Thus, for large enough n, $t < F_n(y)$ which implies $F_n^{-1}(t) \leq y \leq F^{-1}(s) + \varepsilon$. This implies that

$$\limsup_n F_n^{-1}(t) \leq F^{-1}(s) + \varepsilon,$$

for all $s > t$ and $\varepsilon > 0$ since F^{-1} is continuous at t this implies that

$$\limsup_n F_n^{-1}(t) \leq F^{-1}(t).$$

Thus, $F_n^{-1}(t) \to F^{-1}(t)$ as required. $\square$

As a corollary of Skorokhod's theorem we can prove the continuous mapping theorem for convergence in distribution.

Corollary 1. *Suppose that $X_n \stackrel{d}{\to} X$ and g is continuous on a set B with $\mathbb{P}(X_0 \in B) = 1$, then $g(X_n) \stackrel{d}{\to} g(X)$*

Proof. Let X_n^* and X^* be as in Skorokhod's theorem. Then $g(X_n^*) \stackrel{a.s.}{\to} g(X^*)$ be the continuous mapping theorem for almost sure convergence. Thus, $g(X_n^*) \stackrel{d}{\to} g(X^*)$. But $g(X_n^*) \stackrel{\text{dist}}{=} g(X_n)$ and $g(X^*) \stackrel{\text{dist}}{=} g(X)$ and thus $g(X_n) \stackrel{d}{\to} g(X)$. $\square$