Convergence of Random Variables

John Duchi

Stats 300b – Winter Quarter 2021

Outline

- convergence definitions
- continuous mapping and Slutsky's theorems
- big-O notation
- major convergence theorems

Reading: van der Vaart Chapter 2

Basics of convergence

Definition

Let X_n be a sequence of random vectors. Then X_n *converges in probability to* X,

$$X_n \xrightarrow{p} X$$

if for all $\epsilon > 0$,

$$\mathbb{P}(\|X_n - X\| \geq \epsilon) \to 0 \quad \text{as } n \to \infty$$

Convergence in distribution / weak convergence

Definition

For random variables $X_n \in \mathbb{R}$ and $X \in \mathbb{R}$, X_n *converges in distribution to* X,

$$X_n \xrightarrow{d} X \quad \text{or} \quad X_n \rightsquigarrow X$$

if for all x such that $x \mapsto \mathbb{P}(X \leq x)$ is continuous,

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x) \;\; \text{as } n \to \infty$$

Convergence in distribution / weak convergence

Definition

For metric space-valued random variables X_n, X, X_n *converges in distribution to* X if for all bounded continuous f

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \quad \text{as } n \to \infty$$

Convergence in pth mean

Definition

Vector-valued random variables X_n *converge in* L^p (or *converge in pth mean*) to X,

$$X_n \xrightarrow{L^p} X$$

if $\mathbb{E}[\|X_n - X\|^p] \to 0$ as $n \to \infty$.

Almost sure convergence

Definition

Random variables *converge almost surely*, $X_n \stackrel{a.s.}{\to} X$, if

$$\mathbb{P}\left(\lim_n X_n \neq X\right) = 0 \quad \text{or} \quad \mathbb{P}\left(\lim_n \|X_n - X\| \geq \epsilon\right) = 0$$

for all $\epsilon > 0$.

Standard implications

An example (the central limit theorem)

Portmanteau Lemma

Theorem

Let X_n, X be random vectors. The following are all equivalent.

(1) $X_n \xrightarrow{d} X$

(2) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all bounded continuous f*

(3) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all 1-Lipschitz f with $f(x) \in [0, 1]$*

(4) $\liminf_n \mathbb{E}[f(X_n)] \geq \mathbb{E}[f(X)]$ *for all continuous nonnegative 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) $\lim \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$ *for all sets B such that* $\mathbb{P}(X \in \mathrm{bd}B) = 0$

A few remarks on the Portmanteau Lemma

- A collection $\mathcal{F}$ is a *convergence determining class* if

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \text{ for all } f \in \mathcal{F} \quad \text{if and only if} \quad X_n \xrightarrow{d} X$$

- Characteristic functions of the form $f(x) = \exp(it^T x)$ for $i = \sqrt{-1}$ and $t \in \mathbb{R}^d$ are convergence determining

- Boundedness of f in the Portmanteau lemma is important

Proof sketches of the Pormanteau lemma

Continuous mapping theorems

Theorem (Continuous mapping)

Let g be continuous on a set B such that $\mathbb{P}(X \in B) = 1$. Then

(1) $X_n \xrightarrow{p} X$ *implies* $g(X_n) \xrightarrow{p} g(X)$

(2) $X_n \xrightarrow{d} X$ *implies* $g(X_n) \xrightarrow{d} g(X)$

(3) $X_n \xrightarrow{a.s.} X$ *implies* $g(X_n) \xrightarrow{a.s.} g(X)$

Slutsky's theorems

Theorem (Slutsky)

(1) $X_n \xrightarrow{d} c$ *if and only if* $X_n \xrightarrow{p} c$

(2) $X_n \xrightarrow{d} X$ *and* $d(X_n, Y_n) \xrightarrow{p} 0$ *implies that* $Y_n \xrightarrow{d} X$

(3) $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} c$ *implies*

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ c \end{pmatrix}$$

Proof sketches

Consequences of Slutsky's theorems

Corollary

If $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{d} c$*, then*

(1) $X_n + Y_n \xrightarrow{d} X + c$

(2) $X_n Y_n \xrightarrow{d} cX$

(3) *If* $C \in \mathbb{R}^{d \times d}$ *with* $\det(C) \neq 0$*, and* $Y_n \xrightarrow{d} C$*, then* $Y_n^{-1} X_n \xrightarrow{d} C^{-1} X$

Example: t-type statistics

Example

Let $X_i \overset{\text{iid}}{\sim} P$ with $\text{Cov}(X_i) = \Sigma \succ 0$, set

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} X_i \quad S_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_n)(X_i - \mu_n)^T.$$

Then

$$T_n := \frac{1}{\sqrt{n}} S_n^{-1/2} \sum_{i=1}^{n}(X_i - \mu_n) \ \text{ satisfies } \ T_n \overset{d}{\to} \mathcal{N}(0, I).$$

Big-O Notation

Let X_n be random vectors, $R_n \in \mathbb{R}$ be random variables. We write

$$X_n = o_P(R_n) \ \text{ if } \ X_n = Y_n R_n, \quad \text{where } Y_n \xrightarrow{p} 0$$

and

$$X_n = O_P(R_n) \ \text{ if } \ X_n = Y_n R_n \ \text{ where } \ Y_n = O_P(1)$$

and $Y_n = O_P(1)$ means that Y_n is *uniformly tight*, that is,

$$\limsup_{M\to\infty} \sup_n \mathbb{P}\left(\|Y_n\| \geq M\right) = 0.$$

Some consequences of big-O notation

Lemma

We have $o_P(1) + o_P(1) = o_P(1)$, $O_P(1) + O_P(1) = O_P(1)$, *and* $o_P(R_n) = O_P(R_n)$

Lemma

Let $R : \mathbb{R}^d \to \mathbb{R}^k$ *satisfy* $R(0) = 0$ *and let* $X_n \xrightarrow{p} 0$. *Then*

(1) *If* $R(h) = o(\|h\|^p)$ *as* $h \to 0$, *then* $R(X_n) = o_P(\|X_n\|^p)$

(2) *If* $R(h) = O(\|h\|^p)$ *as* $h \to 0$, *then* $R(X_n) = O_P(\|X_n\|^p)$

Compactness and distributional convergence

Definition

A collection of random vectors $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ is *uniformly tight* if for all $\epsilon > 0$, there exists $M < \infty$ such that

$$\sup_{\alpha \in \mathcal{A}} \mathbb{P}\left(\|X_\alpha\| \geq M\right) \leq \epsilon.$$

Remark: A single random vector is tight

Remark: if $X_n \stackrel{d}{\to} X$ then $\{X_n\}_{N \in \mathbb{N}}$ is uniformly tight

Prohorov's Theorem

Theorem (Prohorov)

A collection $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ *is uniformly tight if and only if it is sequentially compact for convergence in distribution, that is, for all sequences* $\{X_n\} \subset \{X_\alpha\}_{\alpha \in \mathcal{A}}$*, there is a subsequence* $n(k)$ *such that* $X_{n(k)} \stackrel{d}{\to} X$ *as* $k \to \infty$ *for some random vector* X*.*

A final useful convergence theorem

Theorem (Scheffe)

For a measure μ, let $f_n \to f$ μ-almost everywhere and assume $\limsup_n \int |f_n|^p d\mu \le \int |f|^p d\mu < \infty$*. Then*

$$\|f_n - f\|_{L^p(\mu)}^p = \int |f_n - f|^p d\mu \to 0$$