Uniform Central Limit Theorems and Convergence in Distribution in Metric Spaces

John Duchi

Stats 300b – Winter Quarter 2021

Outline

- Convergence in distribution in metric spaces
- Compactness in function spaces
- Equi-continuity, finite dimensional convergence, and uniform limits in distribution
- Donsker classes

Reading: van der Vaart, *Asymptotic Statistics*, Chapter 18, 19.1–19.2

Weak convergence

Definition (Convergence in distribution)

Random variables $X_n \stackrel{d}{\to} X$ if

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \text{ for all bounded, continous } f.$$

- definition is the same whether X_n are $\mathbb{R}$-valued or metric-space valued
- sometimes measurability issues for metric-space valued RVs, which we ignore

Tightness

- metric space $(\mathbb{D}, \rho)$

Definition (Tightness)

A $\mathbb{D}$-valued random variable is *tight* if for all $\epsilon > 0$, there exists a compact $K \subset \mathbb{D}$ such that

$$\mathbb{P}(X \in K) \geq 1 - \epsilon.$$

Definition (Asymptotic tightness)

A sequence $X_n \in \mathbb{D}$ of such random variables is *asymptotically tight* if for all $\epsilon > 0$ there exists a compact $K \subset \mathbb{D}$ such that

$$\limsup_{n \to \infty} \mathbb{P}(X_n \notin K^\delta) < \epsilon \quad \text{for all } \delta > 0,$$
$$K^\delta := \{y \in \mathbb{D} \mid \mathrm{dist}(y, K) < \delta\}$$

Prohorov's Theorem

Theorem

Let X_n be $\mathbb{D}$-valued random variables

(i) *If $X_n \xrightarrow{d} X$ where X is a tight random variable, then $\{X_n\}$ is asymptotically tight*

(ii) *If X_n is asymptotically tight, there exists a subsequence $n_k \subset \mathbb{N}$ and a tight $\mathbb{D}$-valued random variable X such that $X_{n_k} \xrightarrow{d} X$*

The idea of Prohorov's theorem

Continuous Functions on Compacta

- (T, d) is a compact metric space
- $L^\infty(T)$ is bounded measurable $f : T \to \mathbb{R}$
- Continuous function $\ell : T \times \mathcal{X} \to \mathbb{R}$, and $X_i \stackrel{\text{iid}}{\sim} P$

The *empirical process* is

$$Z_n(\cdot) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell(\cdot, X_i) - P\ell(\cdot, X) = \sqrt{n}(P_n - P)\ell(\cdot, X)$$

- $Z_n \in L^\infty(T)$
- Z_n is continuous
- For any finite set $t_1, \ldots, t_k$,

$$(Z_n(t_1), \ldots, Z_n(t_k)) \stackrel{d}{\to} \mathcal{N}\left(0, \operatorname{Cov}(\ell(t_i, X), \ell(t_j, X))_{i,j=1}^{k}\right)$$

Compactness in Function Spaces

- Would like to talk about compactness in $L^\infty(T)$
- our limits will be in

$$\mathcal{C}(T, \mathbb{R}) := \{\text{continuous } f : T \to \mathbb{R}\}$$

with metric $\|f - g\|_\infty = \sup_{t \in T} |f(t) - g(t)|$

Arzelà-Ascoli theorem key to compactness in $\mathcal{C}(T, R)$

Uniform continuity

Definition (Modulus of continuity)

For $f : T \to \mathbb{R}$, the *modulus of continuity* of f is

$$\omega_f(\delta) := \sup \{|f(t) - f(s)| : d(s,t) \leq \delta\}$$

Definition (Uniform equicontinuity)

A collection $\mathcal{F} \subset \mathcal{C}(T, \mathbb{R})$ is *uniformly equicontinuous* if

$$\lim_{\delta \downarrow 0} \sup_{f \in \mathcal{F}} \omega_f(\delta) = 0.$$

The Arzelà-Ascoli Theorem

Theorem

Let (T, d) be a compact metric space. The following are equivalent for $\mathcal{F} \subset \mathcal{C}(T, \mathbb{R})$:

(i) *$\mathcal{F}$ is relatively compact (i.e., $\text{cl}\, \mathcal{F}$ is compact in the supremum norm topology)*

(ii) *$\mathcal{F}$ is uniformly equicontinuous and there exists $t_0 \in T$ such that $\sup_{f \in \mathcal{F}} |f(t_0)| < \infty$*

- showing asymptotic tightness will roughly be a stochastic analogue of (ii), uniform equicontinuity

Uniform limits in distribution

Definition (Stochastic equi-continuity)

Let $X_n \in L^\infty(T)$ be random variables. The X_n are *asymptotically stochastically equicontinuous* if for all $\varepsilon, \eta > 0$ there is a partition $T_1, \ldots, T_k$ of T such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\max_i \sup_{s,t\in T_i} |X_{n,s} - X_{n,t}| \geq \varepsilon\right) \leq \eta.$$

Two examples

Example (Linear functions in $\mathbb{R}^d$)

Let $X_i \in \mathbb{R}^d$, $X_i \stackrel{\text{iid}}{\sim} P$ with $P\left\|X\right\|_2^2 < \infty$. Then for $T = \{t \in \mathbb{R}^d \mid \left\|t\right\|_2 \le M\}$, the process $Z_{n,t} := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^\top t$ is asymptotically stochastically equicontinuous.

Example (Generalized linear models)

Let $(X_i, Y_i) \stackrel{\text{iid}}{\sim} P$ and consider losses $\ell(\theta; x, y) = h(\theta^\top x, y)$ for some Lipschitz h with $\mathbb{E}[|h(0, Y)|^2] < \infty$ and $P\left\|X\right\|_2^2 < \infty$, $\theta \in \Theta$ compact. Then $Z_{n,t} := \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell(\theta; X_i, Y_i)$ is A.S.E.

Weak convergence of functions

Theorem

For a process $X_n \in L^\infty(T)$, the following are equivalent.

(i) $X_n \xrightarrow{d} X \in L^\infty(T)$, *where X is tight*

(ii) *we have both*

  (a) *Finite dimensional convergence (FIDI): for all $t_1^k = (t_1, \ldots, t_k) \subset T$, there exists $Z_{t_1^k}$ such that*

$$(X_{n,t_1}, \ldots, X_{n,t_k}) \xrightarrow{d} Z_{t_1^k}$$

  (b) *The X_n are asymptotically stochastically equicontinuous*

1. Constructing a separable subset of T

2. Extending the process to continuous functions

3. Demonstrating convergence proper

Discussion

- showed that limit $\{Z_t\}_{t \in T}$ has uniformly continuous sample paths for a metric ρ on T for which (T, ρ) is a totally bounded metric space

From continuity to a limit distribution

Corollary

Let (T, d) be a totally bounded metric space. Then if

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{d(s,t) \le \delta} |X_{n,s} - X_{n,t}| \ge \varepsilon \right) = 0, \quad \textit{all } \varepsilon > 0,$$

and X_n has finite dimensional convergence, $X_n \xrightarrow{d} X \in L^\infty(T)$ and X is tight

Donsker classes

A collection $\mathcal{F} \subset \mathcal{X} \to \mathbb{R}$ is *P-Donsker* if the processes

$$\left[\sqrt{n}(P_n - P)f\right]_{f \in \mathcal{F}}$$

converge in distribution to a tight limit in $L^\infty(\mathcal{F})$

this limit must be a *Gaussian process* $\mathbb{G} = \mathbb{G}_P \in L^\infty(\mathcal{F})$, i.e., $\mathbb{G}$ is a random mapping $\mathbb{G} : \mathcal{F} \to \mathbb{R}$ with

(i) $\mathbb{E}[\mathbb{G}f] = 0$
(ii) $\mathbb{E}[(\mathbb{G}f)^2] = Pf^2 - (Pf)^2$
(iii) $\mathbb{E}[\mathbb{G}f\mathbb{G}g] = \mathrm{Cov}_P(f, g) = Pfg - (Pf)(Pg)$
(iv) Equivalently to (i)–(iii), for any $f_1, \ldots, f_k \in \mathcal{F}$,

$$(\mathbb{G}f_1, \ldots, \mathbb{G}f_k) \sim \mathcal{N}\left(0, \mathrm{Cov}_P(f_i, f_j)_{i,j=1}^k\right).$$

Brownian bridges

Example (P-Brownian bridge)

For $F_n(t) = P_n(X \le t)$ and $F(t) = P(X \le t)$,

$$\sqrt{n}(F_n(t) - F(t))_{t\in\mathbb{R}} \xrightarrow{d} \mathbb{G}_P$$

in $L^\infty(\mathbb{R})$ and $(\mathbb{G}_t)_{t\in\mathbb{R}}$ has

$$\mathrm{Cov}(\mathbb{G}_t, \mathbb{G}_s) = P(X \le s \wedge t) - P(X \le s)P(X \le t)$$

and Gaussian increments

Limiting Gaussian for Lipschitz losses

Example

Let $\Theta \subset \mathbb{R}^d$ be compact, $\ell : \Theta \times \mathcal{X} \to \mathbb{R}$ be a loss, where $\ell(\cdot, x)$ is $M(x)$-Lipschitz on Θ with $M \in L^2(P)$. Then

$$\sqrt{n}(P_n \ell(\cdot, X) - P\ell(\cdot, X)) \stackrel{d}{\to} \mathbb{G} \in \mathcal{C}(\Theta, \mathbb{R})$$

with $\mathrm{Cov}(\mathbb{G}_{\theta_0}, \mathbb{G}_{\theta_1}) = \mathrm{Cov}(\ell(\theta_0, X), \ell(\theta_1, X))$

Why is this useful?

- let $\widehat{\theta}_n$ be continuous w.r.t. supremum norm on T
- assume $\sqrt{n}(P_n - P) \xrightarrow{d} \mathbb{G}$ in $L^\infty(T)$
- the continuous mapping theorem gives limit distributions of $\widehat{\theta}_n(P_n)$

From entropies to Donsker classes

Theorem

Let $\mathcal{F} \subset \{\mathcal{X} \to \mathbb{R}\}$ have envelope $F : \mathcal{X} \to \mathbb{R}_+$ and assume

$$\int_0^\infty \sup_Q \sqrt{\log N(\mathcal{F}, L^2(Q), (PF^2)^{1/2}\epsilon)} d\epsilon < \infty$$

where the supremum is over finitely supported Q. If $PF^2 < \infty$, then $\mathcal{F}$ is P-Donsker.

Finalizing proof sketch

Examples

Example (VC classes)

If $\mathcal{F}$ is a VC-class with envelope F, then $\log N(\mathcal{F}, L^r(Q), (PF^r)^{1/r}\epsilon) \lesssim r\mathrm{VC}(\mathcal{F}) \log \frac{1}{\epsilon}$, so

$$\sqrt{n}(P_n - P) \xrightarrow{d} \mathbb{G}$$

in $L^\infty(\mathcal{F})$

Example (Brownian bridge)

By above, $\sqrt{n}[F_n(t) - F(t)]_{t\in\mathbb{R}} \xrightarrow{d} \mathbb{G}$, where $\mathrm{Cov}(\mathbb{G}_t, \mathbb{G}_s) = F(t \wedge s) - F(t)F(s)$