Theoretical Statistics, STOR 655

Overview of Weak Convergence and the CLT

Andrew Nobel

January 2023

Review: Families of Continuous Functions

Definition

1. $C_b(\mathbb{R}^d) = \{$bounded continuous functions $f : \mathbb{R}^d \to \mathbb{R}\}$
2. $C_o(\mathbb{R}^d) = \{$continuous $f : \mathbb{R}^d \to \mathbb{R}$ with compact support$\}$

Fact

1. Every $f \in C_o(\mathbb{R}^d)$ is uniformly continuous
2. Every $f \in C_o(\mathbb{R}^d)$ is bounded, so $C_o(\mathbb{R}^d) \subseteq C_b(\mathbb{R}^d)$
3. Every $f \in C_b(\mathbb{R}^d)$ is Borel measurable

Weak Convergence of Random Vectors

Idea: Define the convergence of $X_1, X_2, \ldots$ to a limit X by considering the expectations of smooth functions

Definition: A sequence of random vectors $X_1, X_2, \ldots \in \mathbb{R}^d$ converges weakly (in-distribution) to a random vector X, written $X_n \Rightarrow X$, if

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X) \text{ for every } f \in C_b(\mathbb{R}^d)$$

Important: Random vectors $X_1, X_2, \ldots, X$ can be defined on different probability spaces

Weak Convergence, cont.

Note: If $X_n \sim \mu_n$ then $\mathbb{E}f(X_n) = \int f\, d\mu_n$, so weak convergence depends only on the individual distributions of $X_1, X_2, \ldots$

Definition: A sequence of distributions $\mu_1, \mu_2, \ldots$ on $\mathbb{R}^d$ converges to a distribution μ if $\int f\, d\mu_n \to \int f\, d\mu$ for every $f \in C_b(\mathbb{R}^d)$

Other points

- Definition generalizes to random objects taking values general, possibly infinite dimensional, spaces
- In general, $X_n \Rightarrow X$ does *not* imply that $\mathbb{E}X_n \to \mathbb{E}X$.

Weak Convergence Examples

1. If X_n has the discrete uniform distribution with pmf $p(j/n) = 1/n$ for $j = 1, \ldots, n$ then $X_n \Rightarrow$ Uniform$[0, 1]$

2. If $X_n \sim$ Bin(n, p) for $p \in (0, 1)$ then

$$\frac{X_n - np}{\sqrt{np(1-p)}} \Rightarrow \mathcal{N}(0, 1)$$

3. If μ_n and σ_n^2 are numerical sequences such that $\mu_n \to \mu$ and $\sigma_n^2 \to \sigma^2 > 0$ then $\mathcal{N}(\mu_n, \sigma_n^2) \Rightarrow \mathcal{N}(\mu, \sigma^2)$

Multivariate Central Limit Theorem

Theorem: If $X_1, X_2, \ldots \in \mathbb{R}^d$ are iid with mean $\mathbb{E}(X_i) = \mu$ and finite variance matrix $\mathrm{Var}(X_i) = \Sigma$ then

$$n^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right) \Rightarrow \mathcal{N}_d(0, \Sigma)$$

Note that Σ need not be positive definite.

Many extension of the standard CLT

- Non identically distributed, dependent random vectors
- Random functions, combinatorial structures

Weak Convergence

Theorem: Let $X_1, X_2, \ldots, X \in \mathbb{R}^d$ be random vectors

1. If $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all $f \in C_o(\mathbb{R}^d)$ then $X_n = O_p(1)$
2. If $X_n \Rightarrow X$ then $\{X_n\}$ is stochastically bounded
3. $X_n \Rightarrow X$ iff $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all $f \in C_o(\mathbb{R}^d)$

Corollary: To establish weak convergence, we can replace $C_b(\mathbb{R}^d)$ by the smaller family $C_o(\mathbb{R}^d)$

Weak Convergence and Convergence in Probability

Fact

1. If $X_n \to X$ in probability then $X_n \Rightarrow X$
2. If $X_n \Rightarrow v$ where v is a constant vector, then $X_n \to v$ in probability

Example: If $X \sim \mathcal{N}(0, 1)$ and $X_n = (-1)^n X$ then clearly $X_n \Rightarrow X$, but X_n has no almost sure or in probability limit

Portmanteau Theorem

Theorem: Let $X_1, X_2, \ldots, X \in \mathbb{R}^d$ have CDFs $F_1, F_2, \ldots, F$. TFAE

1. $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all $f \in C_b(\mathbb{R}^d)$
2. $F_n(x) \to F(x)$ for every $x \in \mathbb{R}^d$ where F is continuous
3. $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all $f \in C_o(\mathbb{R}^d)$
4. $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded Lipshitz functions $f : \mathbb{R}^d \to \mathbb{R}$
5. $\mathbb{E}\exp(i\langle X_n, v\rangle) \to \mathbb{E}\exp(i\langle X, v\rangle)$ for all $v \in \mathbb{R}^d$
6. $\limsup_n \mathbb{E}f(X_n) \geq \mathbb{E}f(X)$ for all continuous $f : \mathbb{R}^d \to \mathbb{R}$
7. $\liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ for all open $G \subseteq \mathbb{R}^d$
8. $\limsup_n \mathbb{P}(X_n \in H) \leq \mathbb{P}(X \in H)$ for all closed $H \subseteq \mathbb{R}^d$
9. $\lim_n \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$ for all Borel B s.t. $\mathbb{P}(X \in \overline{B} \setminus B^o) = 0$