STATS300B – Lecture 2

Julia Palacios
Scribed by Michael Howes

01/06/22

Contents

1 Big picture

Suppose we have a model where $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} f_\theta(x)$ where θ is a parameter of interest. Our data $x_1, \ldots, x_n$ is modelled as a realization of $X_1, \ldots, X_n$. We can take this data and compute statistics t_n. By a statistic we just mean any function of the data. Some examples are

$$t_n = \frac{1}{n} \sum_{i=1}^n x_i, \quad t_n = \min\{x_i : 1 \le i \le n\}, \quad t_n = \frac{1}{n} \sum_{i=1}^n x_i^i.$$

We want to understand the properties of our statistics. For example, for estimating $g(\theta)$ we would like to know the bias of a statistic which is defined to be

$$\text{Bias}(T_n) = \mathbb{E}[T_n] - g(\theta).$$

We would also be interested in the asymptotic bias which is $\lim_n \text{Bias}(T_n)$. Similarly, we could ask if our statistics are consistent for $g(\theta)$. That is does $T_n \overset{p}{\to} g(\theta)$ for all θ. Consistency means that $T_n - \theta \overset{p}{\to} 0$, but consistency does not tell us about the rate at which this convergence happens. To understand the rate, we need to talk about convergence in distribution. We will see that another desirable property of a sequence of statistics is that the sequence asymptotically normal. That is, for some σ^2,

$$\sqrt{n}(T_n - g(\theta)) \overset{d}{\to} \mathsf{N}(0, \sigma^2).$$

2 An example

Before we discuss convergence in distribution, let's look at an example. Suppose $X_1, \ldots, X_n$ i.i.d. random variables with mean μ and variance σ^2. Suppose both μ and σ^2 are unknown, and we are interested in estimating σ^2. Consider

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Note that $S_n^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \right)$. By the weak law of large numbers we know that

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{p} \mathbb{E}[X^2] = \mu^2 + \sigma^2,$$

and

$$\bar{X}_n \xrightarrow{p} \mathbb{E}[X] = \mu.$$

Thus, by the continuity theorem $\bar{X}_n^2 \xrightarrow{p} \mu^2$ and thus

$$S_n^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \right) \xrightarrow{p} \sigma^2.$$

But, does $\sqrt{n}(S_n^2 - \sigma^2)$ have a "limiting distribution."

3   Convergence in distribution

Definition 1. A sequence of random vectors X_n *converge in distribution* to X if for all x such that $x \mapsto \mathbb{P}(X \le x)$ is continuous at x,

$$\mathbb{P}(X_n \le x) \to \mathbb{P}(X \le x).$$

We will denote convergence in distribution by $X_n \xrightarrow{d} X$.

Remark 1. For vectors $x, y \in \mathbb{R}^d$, $x \le y$ means that $x_j \le y_j$ for all $1 \le j \le d$.

Example 1. Suppose $X_n \sim \mathsf{Exp}(\lambda_n)$ where $\lambda_n = \frac{n}{n+1}$. Then

$$\mathbb{P}(X_n \le t) = \begin{cases} 0 & \text{if } t \le 0, \\ 1 - e^{-\frac{n}{n+1}t} & \text{if } t > 0. \end{cases}$$

Which converges to

$$F(t) = \begin{cases} 0 & \text{if } t \le 0, \\ 1 - e^{-t} & \text{if } t > 0. \end{cases}$$

Thus, $X_n \xrightarrow{d} X$ where $X \sim \mathsf{Exp}(1)$.

Example 2. The condition that x is a point of continuity of $x \mapsto \mathbb{P}(X \le x)$ is important. Consider $Y_n \sim \mathsf{N}(0, 1/n)$. Then, by a change of variables,

$$F_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi n}} e^{-n\omega^2/2} d\omega = \int_{-\infty}^{x\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Thus, as $n \to \infty$, $F_n(x)$ converges to $G(x)$ where

$$G(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{2} & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

The CDF of $X \equiv 0$ is $F(x) = \mathbf{1}_{[0,\infty)}(x)$ which is discontinuous at 0 and agrees with $G(x)$ at all other points. Thus, $X_n \xrightarrow{d} X$ even thought $\lim_n F_n(x) \ne F(0)$.

Remark 2. The dominated convergence theorem is an important result about exchanging integrals and limits. It should be reviewed if you are not familiar with it.

There are multiple characterizations of convergence in distribution. One if the Helly–Brag Theorem.

Lemma 1. *A sequence of random vectors X_n converges in distribution to X if and only if for all bounded and continuous f, $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(x)]$.*

The above condition is sometimes taken as the definition of convergence in distribution. The above condition generalizes more easily to arbitrary metric spaces. One can also replace the collection of all bounded and continuous functions with a smaller class of functions such as the Lipschitz functions or the characteristic functions $f(x) = e^{itx}$ for $t \in \mathbb{R}$. A class of functions with the above property is called a class of convergence determining functions.

The main result about convergence in distribution is the central limit theorem.

Theorem 1. *Suppose $X_1, X_2, \ldots$ are i.i.d. with mean μ and variance $\sigma^2 < \infty$, then*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathsf{N}(0, \sigma^2).$$

The central limit theorem is more informative than the weak law of large numbers (but it has stronger assumptions). There is also a multivariate version of the central limit theorem.

Theorem 2. *Suppose $X_1, X_2, \ldots$ are i.i.d. random vectors in $\mathbb{R}^d$ with mean $\mathbb{E}[X_1] = \mu$ and covariance matrix $\Sigma = \mathbb{E}(X_1 - \mu)(X_1 - \mu)^T$ with finite entries. Then*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathsf{N}_d(0, \Sigma).$$

Two other versions of the central limit theorem are:

Theorem 3 (Lyapunov's CLT). *Fix $\delta > 0$. Suppose that for each n, $X_{n,1}, \ldots, X_{n,n}$ are independent random variables with means $\mu_{n,i} = \mathbb{E}[X_{n,i}]$ and variance $\sigma_{n,i}^2 = \mathrm{Var}(X_{n,i})$. Suppose also that the moments $\gamma_{n,i} = \mathbb{E}|X_{n,i} - \mu_{n,i}|^{2+\delta}$ are all finite. Let $\mu_n = \sum_{i=1}^n \mu_{n,i}$ and $\sigma_n^2 = \sum_{i=1}^n \sigma_{n,i}^2$ and let $\gamma_n = \sum_{i=1}^n \gamma_{n,i}$. If $\frac{\gamma_n}{\sigma_n^{2+\delta}} \to 0$, then*

$$\frac{1}{\sigma_n} \sum_{i=1}^n X_{n,i} - \mu_{n,i} \to \mathsf{N}(0, 1).$$

The version of the CLT proved in Persi's STATS 310A:

Theorem 4 (Lindeberg–Feller CLT). *Fix $\delta > 0$. Suppose that for each n, $X_{n,1}, \ldots, X_{n,n}$ are independent random variables with mean 0 and variance $\sigma_{n,i}^2 = \mathrm{Var}(X_{n,i})$. Let $\sigma_n^2 = \sum_{i=1}^n \sigma_{n,i}^2$ and $S_n = \sum_{i=1}^n X_{n,i}$. Then both $\frac{S_n}{\sigma} \xrightarrow{d} \mathsf{N}(0, 1)$ and $\max\{\sigma_{n,i}^2/\sigma_n^2 : 1 \le i \le n\} \to 0$, if and only if $X_{n,i}$ satisfy* Lindeberg's condition. *That is for all $\varepsilon > 0$,*

$$\frac{1}{\sigma_n^2} \sum_{i=1}^n \mathbb{E}[X_{n,i}^2 \mathbf{1}_{|X_{n,i}| > \varepsilon \sigma_n}] \to 0.$$

Another type of convergence is convergence in L^p.

Definition 2. For $p \in [1, \infty]$, A sequence of random vectors X_n is said to *converge in L^p* to the random vector X if $\mathbb{E}[\|X_n - X\|^p] \to 0$, where $\|\cdot\|$ is any norm on $\mathbb{R}^d$. We denote convergence in L^p by $X_n \xrightarrow{L^p} X$.

Note that by Markov's inequality, convergence in L^p for any p implies convergence in probability. Convergence in L^p also implies convergence in $L^{p'}$ for any $p' \leq p$.

As stated before there are many equivalent ways to describe convergence in distribution. Some of them are summarized by the bellow theorem which is called the Portmantaeu theorem.

Theorem 5. *Let X_n and X be a random vectors. The following are equivalent.*

1. $X_n \xrightarrow{d} X$.

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.*