# STATS300B – Lecture 1

Julia Palacios
Scribed by Michael Howes

01/04/22

## Contents



# 1 Overview

This course is about asymptotic statistics and what happens when the number of samples goes to infinity. We'll cover three main topics

1. Finite dimensional problems and models.
2. Optimality and comparisons.
3. Infinite dimensional and uniform problems.

This course is useful for statistics, machine learning, computer science and data science. The marks for the course will be calculated as follows:

- Weekly homework (40%),
- Midterm exam (30%),
- Final exam (25%),
- Participation (5%).

The midterm will take place on February 8 and will be in class. There are three textbooks for the class

- Asymptotic statistics by van der Vaart (this will be the main text).
- High – dimensional probability by Vershynin (this will be used towards the end of the course).
- Theoretical statistics by Keener (this will be used a bit at the start of the course).

## 2 Convergence of random variables

See van der Vaart Ch 2,3 and Keener Ch 8.

In STATS 300A we studied properties of estimators that held in an *exact* sense (for example unbiased, MRE, UMRUE, etc.). In this course we will study properties that hold "approximately" or "in the limit." To make this precise we have to discuss what we mean by taking limits of estimators and limits of random variables.

**Definition 1.** Let $X_1, X_2, \ldots$ and $X$ be random vectors. We say that $X_n$ *converges to* $X$ *in probability* and write $X_n \xrightarrow{p} X$ if for every $\varepsilon > 0$,

$$\mathbb{P}(\|X_n - X\| > \varepsilon) \to 0,$$

as $n \to \infty$.

Note that be replacing $\|X_n - X\|$ with $d(X_n, X)$ we can generalize convergence in probability to arbitrary metric spaces.

**Theorem 1** (Chebyshev's inequality)**.** *For any random variable $X$ and any constant $a > 0$,*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[X^2]}{a^2}.$$

*Proof.* Note that $\mathbf{1}_{|X| \geq a} \leq \frac{X^2}{a^2}$. Thus, by monotonicity and linearity of expectation we have

$$\mathbb{P}(|X| \geq a) = \mathbb{E}[\mathbf{1}_{|X|\geq a}] \leq \mathbb{E}\left[\frac{X^2}{a^2}\right] = \frac{\mathbb{E}[X^2]}{a^2}.$$ □

Chebyshev's inequality is a powerful tool for proving convergence in probability.

**Proposition 1.** *If $\mathbb{E}(Y_n - Y)^2 \to 0$ and $n \to \infty$, then $Y_n \xrightarrow{p} Y$.*

*Proof.* Note that for all $\varepsilon > 0$,

$$0 \leq \mathbb{P}(|Y_n - Y| \geq \varepsilon) \leq \frac{\mathbb{E}(Y_n - Y)^2}{\varepsilon^2}.$$

Thus, $\lim_{n\to\infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon) \to 0.$ □

**Example 1.** Suppose $X_1, X_2, \ldots$ are i.i.d. with mean $\mu$ and variance $\sigma^2$, then $\bar{X}_n \xrightarrow{p} \mu$.

*Proof.* Note that,

$$\mathbb{E}(\bar{X}_n - \mu)^2 = \mathrm{Var}(\bar{X}_n) = \frac{\mathrm{Var}(X_1)}{n} = \frac{\sigma^2}{n}.$$

So $\mathbb{E}(\bar{X}_n - \mu)^2 \to 0.$ □

The next definition relates convergence in probability to estimation.

**Definition 2.** A sequence of estimators $\delta_n$ is consistent for $g(\theta)$ if for all $\theta \in \Omega$, $\delta_n \xrightarrow{p} g(\theta)$ as $n \to \infty$.

**Remark 1.** When using squared error loss, we consider the MSE $R(\theta, \delta_n) = \mathbb{E}_\theta(\delta_n - g(\theta))^2$. By Proposition 1, if the MSE goes to 0 for all $\theta$, then $\delta_n$ is consistent. Recall that we have the decomposition

$$R(\theta, \delta_n) = b_n^2(\theta) + \mathrm{Var}_\theta(\delta_n),$$

where $b_n(\theta)$ is the bias of $\delta_n$. Thus, if the bias and variance of a sequence of estimators go to zero, then the estimators are consistent.

**Exercise 1.** If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $X_n + Y_n \xrightarrow{p} X + Y$.

*Proof.* Recall that

$$\|X_n + Y_n - (X+Y)\| \leq \|X_n - X\| + \|Y_n - Y\|.$$

Thus,

$$\mathbb{P}(\|X_n + Y_n - (X+Y)\| \geq \varepsilon) \leq \mathbb{P}(\|X_n - X\| + \|Y_n - Y\| \geq \varepsilon).$$

Also, if $\|X_n - X\| + \|Y_n - Y\| \geq \varepsilon$, then $\|X_n - X\| \geq \varepsilon/2$ or $\|Y_n - Y\| \geq \varepsilon/2$. Thus,

$$\mathbb{P}(\|X_n + Y_n - (X+Y)\| \geq \varepsilon) \leq \mathbb{P}(\|X_n - X\| \geq \varepsilon/2) + \mathbb{P}(\|Y_n - Y\| \geq \varepsilon/2) \to 0,$$

as $n \to \infty$. $\square$

**Exercise 2.** If $X_n \xrightarrow{p} a$ where $a$ is a constant and $g$ is a function that is continuous at $a$, then $g(X_n) \xrightarrow{p} g(a)$.

*Proof.* Let $\varepsilon > 0$. There exists $\delta > 0$ such that $\|x - a\| < \delta$ implies $\|g(x) - g(a)\| < \varepsilon$. Thus,

$$\mathbb{P}(\|g(X_n) - g(a)\| \geq \varepsilon) \leq \mathbb{P}(\|X_n - a\| \geq \delta) \to 0,$$

as $n \to \infty$. $\square$