

Principles of Statistics

ARTHUR CONMY*

Part II, Michaelmas Term 2021

These notes are based on lectures given (virtually) by Dr P-L Loh in Michaelmas term 2021. Credit is also due to Evan Chen for the style file for these notes¹.

Contents

1	Introduction	1
2	Convergence	1

§1 Introduction

(skip if not interested in meta notes)

This course follows naturally on from IB Statistics. It also borrows some technology from II Probability and Measure. I haven't currently decided what emphasis to add to these notes: a focus on intuitions or rigour or something else entirely.

§2 Convergence

Since we're in a non-deterministic setting, we need new definitions of convergence. Let $X, X_1, X_2, \dots : \mathcal{X} \rightarrow \mathbb{R}$ be random variables, where $\mathcal{X} \subseteq \mathbb{R}^p$. The first and weakest notion of convergence is *convergence in distribution*.

Definition 2.1 (Convergence in distribution). $X_n \xrightarrow{d} X$ if

$$\mathbb{P}[X_n \leq t] \rightarrow \mathbb{P}[X \leq t] \quad (1)$$

for all $t \in \mathbb{R}^p$ for which $t \rightarrow \mathbb{P}[X \leq t]$ is continuous.

Here, \leq means $x_i \leq y_i$ in all components.

A stronger notion is convergence in probability.

Definition 2.2 (Convergence in probability). $X_n \xrightarrow{P} X$ if

$$\lim_{n \rightarrow \infty} \mathbb{P}[\|X_n - X\| > \varepsilon] = 0. \quad (2)$$

The proof that convergence in probability implies convergence in distribution is omitted.

*Please send any corrections and/or feedback to asc70@cam.ac.uk

¹Available here: <https://github.com/vEnhance/dotfiles/blob/master/texmf/tex/latex/evan/evan.sty>.

Example 2.3 (Convergence in distribution $\not\Rightarrow$ convergence in probability)

Consider X, X_1, X_2, \dots independent and identically distributed $\mathcal{N}(0, 1)$ random variables. Then all these have identical distribution function, but certainly $X_n \xrightarrow{P} X$.

The strongest notion of convergence is convergence *almost surely*. This is closely related to null sets from measure theory, i.e not necessarily empty $A \in \mathfrak{B}$ with $\mu(A) = 0$.

Definition 2.4 (Convergence almost surely). $X_n \xrightarrow{a.s} X$ if

$$\mathbb{P}[||X_n - X|| \rightarrow 0] = 1. \quad (3)$$

Example 2.5 (Convergence in probability $\not\Rightarrow$ convergence in distribution)

Consider independent X_1, X_2, \dots with $X_n \sim \text{Ber}(\frac{n-1}{n})^a$, and X the constant random variable 1 (equivalently $X_n \sim \text{Ber}(1)$). Then (check!) $X_n \xrightarrow{P} X$ but $X_n \not\xrightarrow{a.s} X$.

^aThese random variables can be interpreted as the boolean outcomes of whether an ever-improving skeleton archer hits a bullseye on the n th shot at a target.

References

- [1] Rajen D. Shah (2021), *Mathematics of Machine Learning*, http://www.statslab.cam.ac.uk/~rds37/teaching/machine_learning/notes.pdf.
- [2] Philippe Rigollet, *18.657: Mathematics of Machine Learning*, https://ocw.mit.edu/courses/mathematics/18-657-mathematics-of-machine-learning-fall-2015/lecture-notes/MIT18_657F15_LecNote.pdf.