<u>36-705 – Intermediate Statistics</u>

<u>Key areas to understand</u>

A pedagogical tool to track the key aspects of the lecture that the instructor emphasises, and as a checklist of things you have judged are important.

List what you feel a need to commit to memory later.

<u>Week 1</u>

<u>Lecture Notes 1 – Review of probability</u>

<u>Youtube lecture 31/08/2016</u>

- There was a lecture that covered earlier parts of the probability review that is not covered here.
- Understand that independence can be a physical fact, or an assumption that has to be evaluated.
- Distinction between parameter and random variable.
- The importance of the IID data to the statistical setting.
- Distinction between the distribution of IID random variables, and the (sampling) distributions of statistics/estimators of IID random variables.
- Source of stochasticity in sampling estimators, and their characterisation as random variables/deterministic functions of random variables.
- Properties of the mean and variance of sample mean and variance estimators i.e. properties of the sampling distribution, and their relation to the mean and variance of the underling IID random variables.

<u>Lecture Notes 2 – Inequalities</u>

<u>Youtube lecture 31/08/2016</u>

- The proof of Lemma 4 (as part of a strategy to prove Hoeffding) is not in the video footage.
- Appreciate role of probability bounds in machine learning, particularly Hoeffding's inequality.
- Probability bounds -> VC theory -> convergence bounds.
- Understand the proofs of the Gaussian tail, Markov, Chebyshev inequalities.

<u>Youtube lecture 02/09/2016</u>

- Understand the proofs of Hoeffding's lemma and Chernoff's method.
- Understand the proof of Hoeffding's inequality, the variational trick.
- Understand role of Cauchy-Schwarz and Jensen's inequality for to bound expectations.
- Application to Kullback-Leibler divergence as having certain desirable distance metric properties.
- Understand the proof of the bound on the expectation of the maximum of a series of IID random variables.
- Appreciate the relation between a "thin-tail" in a probability distribution and an exponential bound on the corresponding moment-generating function – as a sub-Gaussian random variable.

<u>Week 2</u>

<u>Youtube lecture 07/09/2016</u>

- Understand the computer-science definitions of little-o and Big-O notation.
- Understand the adaptation of these definitions into a statistical/probabilistic context i.e. little-op and Big-Op.
- Understand that they are both notions of convergence in probability and stochastic boundedness.
- Understand and be able to use arguments from mathematical analysis and probability inequalities to prove composite statements about random variables stipulated in this notation.

<u>Lecture Notes 3 – Uniform bounds</u>

- Appreciate the limitations of Markov's, Chebyshev's and Hoeffding's inequality as bounds on one random variable.
- Appreciate that many applications in statistics and machine learning require bounds on multiple random variables.
- Understand the definition of the empirical CDF and theoretical CDF.
- Understand the distinction between pointwise and uniform convergence.
- Understand this distinction in context of estimation error on CDF and of training error in classification.

Week 3

<u>Youtube lecture 12/09/2016</u>

- Understand the statement of uniform bounds over finite classes of sets, or of functions , and its derivation.
- Understand the role of the complexity of the class of sets.
- Understand shattering, and the shatter coefficient.
- Understand how to apply these concepts to familiar classes of sets.
- Understand the theorem due to Vapnik-Chervonenkis giving a uniform bound over a class of (possibly infinite) sets in terms of an exponential and shattering coefficient.
- Understand the relation between the shattering coefficient and VC dimension.
- Understand Sauer's theorem on the behaviour of shattering coefficient and its relation with the VC dimension.

<u>Lecture Notes 4 – Convergence theory</u>

- Understand the notion of a statistic.
- Understand the distinction between properties of a sequence of random variables, and properties of a sequence of statistics.
- Appreciate the distinction between convergence in the setting of mathematical analysis, and in the probabilistic setting.
- Understand the following probabilistic formulations of convergence: almost sure convergence, convergence in probability, convergence in quadratic mean, and convergence in distribution.
- Understand that the main preoccupations of the course are convergence in probability, and in distribution.
- Understand the taxonomy of relationships between these formulations of convergence.
- Understand how to prove convergence in distribution.

<u>Youtube lecture 14/09/2016</u>

- Understand how to prove convergence in probability.

- Understand the directions of implication in the taxonomy of convergences.
- Understand the proof for convergence in quadratic mean implying convergence in probability.
- Understand the proof for convergence in probability implying convergence in distribution.
- Understand the proof that convergence in probability does not imply convergence in quadratic mean, by counterexample.
- Understand that convergence in quadratic mean is a relatively stronger statement about the moments and tails of a distribution.
- Understand the proof that convergence in distribution does not imply convergence in probability.
- Understand that convergence in probability is a statement about random variables, whilst convergence in distribution is about the probability statements one makes about random variables.
- Understand how convergence in distribution can be combined with rescaling to give meaningful approximations in situations where working with the existing probability distribution is intractable.
- Understand the theorems concerning preservation of convergence under transformations, Slutsky's theorem, and the continuous mapping theorem.
- Understand that the LLN and CLT are related to convergence in probability and distribution respectively.
- Understand the statement and proof of the Weak Law of Large Numbers (WLLN) for finite variance.
- Understand the statement of the Strong Law of Large Numbers.