36-771 Fall 2018

Martingales 1 : Concentration inequalities

Classical asymptotic and nonasymptotic results

Lecturer : Aaditya Ramdas

1 Classical limit theorems

In the following, we will always denote $S_n := X_1 + \cdots + X_n$.

Theorem 1 (Strong Law of Large Numbers (SLLN)) *Let X_i be iid with mean μ. Then*

$$\frac{S_n}{n} \to \mu, \text{ almost surely.}$$

What's so strong about the strong law? The weak law states that $S_n/n \to \mu$ in probability, which is a weaker statement. So can we just forget about the weak law, and only study the strong law?! No, because the weak law actually holds under weaker assumptions. Despite what Wikipedia, Wolfram and other websites currently say, here is a more complete description of the WLLN.

Theorem 2 (Weak Law of Large Numbers (WLLN) from Feller 1971, page 565) *Let X_i be iid with characteristic function ϕ and CDF F. Then the following three conditions are equivalent:*

1. *ϕ is differentiable at 0, and $\phi'(0) = i\mu$.*
2. *As $t \to \infty$, we have $t[1 - F(t) + F(-t)] \to 0$ and $\int_{-t}^{t} xF(dx) \to \mu$.*
3. *$S_n/n \to \mu$, in probability.*

Example 3 (from Charles Geyer's lecture notes at UMN) *Define a random variable via its CDF:*

$$F(t) = \begin{cases} 1 - \frac{\log 2}{t \log t}, t \geq 2 \\ 1/2, -2 \leq t \leq 2 \\ \frac{\log 2}{|t| \log |t|}, t \leq -2 \end{cases}$$

Then one can show that its mean does not exist, and hence by Theorem 4(c) in Ferguson (A Course in Large Sample Theory, 1996), the SLLN does not hold. However, the above random variable is symmetric by construction, and condition (ii) above can be verified to hold with $\mu = 0$. Hence, the third condition (WLLN) holds.

However these theorems do not provide a rate of convergence of sample averages to μ, so this is not sufficient. However, the CLT provides an *asymptotic* rate of convergence, and hence an asymptotic confidence interval.

Theorem 4 (Central Limit Theorem (CLT)) *Let X_i be iid with mean μ and variance σ^2. Then, we have*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1), \text{ in distribution.}$$

Hence, if one divides S_n by n it is damped to zero, if one divides by $\sqrt{n}$, it can still be unbounded, and it is interesting to ask what function of n one needs to divide by in order to get a nontrivial bound. It turns out that dividing S_n by $n^{1/2+\epsilon}$, for any $\epsilon > 0$, results in a limit of zero. Hence, the "right" quantity has to be larger than $n^{1/2}$ but smaller than $n^{1/2+\epsilon}$ for any constant $\epsilon > 0$. However, we have:

Theorem 5 (Law of the iterated logarithm (LIL)) *Let X_i be a symmetric Rademacher (± 1 with equal probability). Then*

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{n \log\log n}} = \sqrt{2}, \text{ a.s.}$$

2 Nonasymptotic bounds

In order to get something non-asymptotic from the CLT, one needs strictly more than two moments, as exemplified by the following Berry-Esseen bound.

Theorem 6 (Berry-Esseen theorem) *Let X_i be iid with mean μ, variance σ^2, and $\mathbb{E}|X_i|^3 < \infty$. Then, we have*

$$\sup_t \left| P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) - \Phi(t) \right| \leq \frac{C\mathbb{E}|X_i|^3}{\sigma^{3/2}\sqrt{n}}.$$

where $0.41 \leq C \leq 0.4748$ is a universal constant.

If we assume more, such as the random variable being bounded, or having a bounded MGF, we can prove many "tail" inequalities, such as Hoeffding's, Bennett's and Bernstein's inequalities. As an example, consider Hoeffding's inequality for bounded random variables.

Theorem 7 (Hoeffding's inequality) *Let X_i be iid, mean 0, bounded in $[-a_i, a_i]$. Denoting $A^2 := \sum_{i=1}^n a_i^2/n$, then*

$$P\left(\frac{S_n}{n} > \epsilon\right) \leq \exp(-2n\epsilon^2/A^2).$$

There are also matrix extensions of these inequalities due to Ahlswede-Winter, Vershynin, Tropp and others.

3 The rest of this mini

The rest of this mini will be devoted to understanding one assumption, and one theorem. The assumption is a "canonical supermartingale assumption", and it is weaker than many standard nonparametric assumptions in the literature. The theorem is informally called the "mother of all exponential concentration inequalities", and it is stronger than many standard famous named theorems in the literaure.

The word "stronger" will become clearer in the rest of the course. For this, we introduce the A-B-C-D-E mnemonics. A: weaker assumptions, B: lower boundary, C: continuous time, D: higher dimensions, E: larger exponent. The meaning of those terms will become clearer later in the course.

Remarkably, we can even improve the aforementioned popular Hoeffding's inequality: it will hold under weaker dependence assumptions, have a lower boundary, have a continuous-time extension, extend to hold for matrices, and also have a tighter exponent (specifically holding when $A = \sum_{i=1}^n a_i^2/2n + \sum_{i=1}^n \mathbb{E}X_i^2/2n$).

We will encounter "self-normalized" ineqaulities for heavy-tailed distributions, concentration for matrices, and if there is time, concentration of continuous time processes, and martingales in smooth Banach spaces.

This course will require the use of (super)martingales, filtrations, convex analysis, and linear algebra. We will revise some of this background next, but most of it is assumed to be known (prerequisites).