E2 207: Concentration Inequalities

Naman Mishra

August 2025

Contents

Lectures

The course

Resources:

Lecture 1. Tuesday August 5

(1)

Evaluation:

-

Question .1. *Toss a fair coin* 10000 *times. How many heads do you expect to see?*

Our expectation is 5000.

Theorem .2 (Weak law of large numbers)**.** *If $X_1, X_2, \ldots$ are iid random variables with finite mean and variance, then $\frac{1}{n}\sum_{i=1}^n X_n \xrightarrow{\mathsf{P}} \mathbf{E}[X_1]$. That is, for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbf{E}[X_1]\right| > \epsilon\right\} = 0.$$

The finite variance assumption is not necessary.

For a fixed ε, the error bound on S_n is $O(n)$.

Theorem .3 (Central limit theorem)**.** *If $X_1, X_2, \ldots$ are iid random variables with finite mean μ and variance σ, then*

$$\frac{1}{\sqrt{n}}(\sum_{i=1}^n (X_i - \mu)) \xrightarrow{d} N(0, \sigma^2).$$

Using this, once more, we have that the number of heads is $5000 \pm \sqrt{10000(\frac{1}{4})}Q^{-1}(0.005)$ with probability at least 0.99, if we assume that $10000 = \infty$, where $Q(x) = \mathbf{P}(Z \geq x)$ for $Z \sim N(0,1)$. $Q^{-1}(0.005) \approx \sqrt{(} - \log 0.005)$. The uncertainty here happens to be around 115 tosses.

We will eventually see the Berry-Esseen theorem.

Theorem .4 (Berry-Esseen inequality)**.** *If $X_1, X_2, \ldots$ are iid with zero mean and finite* ***third*** *moment, then there exists a constant C such that for each n and ε,*

$$\left| \mathbf{P}\left\{ \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i > \varepsilon \right\} - Q(\varepsilon) \right| \leq \frac{C}{\sqrt{n}},$$

The constant here is known to be between 0.4 and 0.5.

Theorem .5 (Large deviations)**.** *For every interval $A \subseteq \mathbb{R}$,*

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbf{P}\left\{ \frac{1}{n} \sum_{i=1}^{n} X_i \in A \right\} = \inf_{x \in A} I(x),$$

where $I(\cdot)$ is the rate function*.*

Thus assuming $10000 = \infty$, we can say that for $x > \frac{1}{2}$,

$$\mathbf{P}(S_n > nx) = \exp(-nI(x)).$$

We have written $I(x)$ in place of $\inf_{y>x} I(y)$ because I happens to be decreasing in this case.

These were largely asymptoptic results. The two simplest concentration inequalities are non-asymptotic.

Theorem .6 (Markov's inequality)**.** *If X is a non-negative random variable, then for any $t > 0$,*

$$\mathbf{P}(X \geq t) \leq \frac{\mathbf{E}[X]}{t}.$$

Theorem .7 (Chebyshev's inequality)**.** *If X is a random variable with finite variance, then for any $\delta > 0$,*

$$\mathbf{P}(|X - \mu| \geq \delta) \leq \frac{\sigma^2}{\delta^2}.$$

Chebyshev's inequaity gives that the number of heads is $5000 \pm \sqrt{\frac{2500}{0.01}}$ with probability at least 0.99. (The uncertainty is 500 tosses.)

In general, CLT and Chebyshev's inequality give results

$$n\,\mathbf{E}\,X_1 \pm \sqrt{n \operatorname{Var} X_1}\sqrt{\log \frac{2}{\delta}},$$

$$\text{and } n\,\mathbf{E}\,X_1 \pm \sqrt{n \operatorname{Var} X_1}\sqrt{\frac{1}{\delta}}$$

with probability at least $1 - \delta$, repectively.

The Chernoff bound, for iid Bernoulli random variables, gives

$$n\,\mathbf{E}\,X_1 \pm C\sqrt{n\operatorname{Var}X_1}\sqrt{\log\frac{1}{\delta}}$$

for some constant C.

Roadmap

We'll study Chernoff and Cramér bounds, which lead to Hoeffding's, Bennett's and Bernstein's inequalities.

We'll look at the coupon collector's problem.

Another question with applications in bioinformatics and `git diff`ing is the following.

Question .8. *Given two iid finite sequences $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, what is the length of a longest common subsequence?*

A common subsequence *of length k is a sequence of length k that is a (not necessarily contiguous) subsequence of both X and Y.*

Theorem .9 (Talagrand's principle)**.** *A random variable that depends smoothly on a large number of random variables satisfies Chernoff-type bounds.*

Methods

- Chernoff-type bounds
- Tensorization techniques to break a function of random variables into its constituents using the (1) martingale, (2) Effron-Stein, and (3) entropy methods.
- Applying isoperimetric inequalities to probability spaces.
- Transportation method

Applications

- Johnson-Lindenstrauss lemma
- Hypercontractivity
- Bounds on the performance of LASSO
- Blowing-up lemmas
- Bin-packing, coupon collector, birthday paradox, LCS, et cetera.

Lecture 1