EE 53100 Concentration Inequalities

Introduction and preliminaries

- Course webpage:
  https://people.iith.ac.in/shashankvatedka/html/courses/2023/EE5603/course_details.html

- Announcements, homework submissions: Google classroom
  (send me an email if you do not get an invite by tomorrow)

- Prerequisites:
  - Strong foundation in probability and random processes
  - Some background in information theory/statistics/machine learning is helpful, but not mandatory
  - Programming in python

- Class timings: Slot B (Mon 10am, Tue 9am, Thu 11am)

- Assessment:
  - homeworks (roughly 3, totaling 55%)
  - 3 quizzes/tests (15% each)

- References: See course webpage

Motivation and background

1. n tosses of a fair coin

For large n, fraction of heads $\approx \frac{1}{2}$

What about finite n?

$\Pr[\text{ fraction of heads} \geq \frac{1}{2} + \delta]$?

$$\Pr[\#\text{ heads} = k] = \binom{n}{k} \left(\frac{1}{2}\right)^n$$

$$\Pr\left[\#\text{ heads} \geq n\left(\frac{1}{2} + \delta\right)\right] = \quad ?$$

② Communication systems :

$$Y = X + Z$$

$$Z \sim \mathcal{N}(0, \sigma^2)$$

(i) $X \in \{+\sqrt{P}, -\sqrt{P}\}$

$$\hat{X} = \begin{cases} +\sqrt{P} & \text{if } Y > 0 \\ -\sqrt{P} & \text{if } Y \leq 0 \end{cases}$$

What is $\Pr[\hat{X} \neq X]$?

(ii) $Y^n = X^n + Z^n$ $\quad X^n \in \mathbb{R}^n$ $\quad Z^n \sim \text{iid } \mathcal{N}(0, \sigma^2)$

$X^n \in C$ codebook

$$\hat{X}^n = \underset{x^n \in C}{\operatorname{argmin}} \|Y^n - x^n\|^2$$

What is $\Pr[\hat{X}^n \neq X^n]$?

③ Empirical risk minimization:

Binary classification: $(X,Y) \sim P_{XY}$ $\quad X \in \mathcal{X}$

$Y \in \{0,1\}$

- Do not know P_{XY}
- Have access to $\underbrace{(X_1,Y_1)(X_2,Y_2) \cdots (X_n,Y_n)}_{\text{dataset}}$ iid P_{XY}

- Problem: Choose $f \in \mathcal{F}$ that minimizes expected risk

$$R(P_{XY}) = \mathbb{E}\left[L(f(X),Y)\right]$$

eg: $$R(P_{XY}) = \mathbb{E}\left[\mathbb{1}_{\{f(X) \neq Y\}}\right] = \Pr\left[f(X) \neq Y\right]$$

- Choose $f : \mathcal{X} \to \{0,1\}$ that best approximates data

$$\hat{f} = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \ \frac{1}{n} \sum_{i=1}^{n} L\left(f(x_i), y_i \right)$$

- How well does this approximate $\mathbb{E}\left[L(f(X), Y) \right]$?

Probability Basics

① Probability space $(\Omega, \mathcal{F}, P)$

Ω - Sample space (collection of outcomes)

$\mathcal{F}$ - Event space

P - Probability measure

Event space : Should form a sigma algebra

① $\Omega \in \mathcal{F}$

② $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

③ Countable collection $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Examples :

① $\Omega = \{1, 2, 3, 4, 5, 6\}$

What is the smallest event space? $\{\emptyset, \Omega\}$

What is the smallest event space containing $\{\{1,2\}, \{2\}\}$?

What is the largest event space? $\{\emptyset, \Omega, \{1,2\}, \{2\}, \{3,4,5,6\}, \{1,3,4,5,6\}, \{2,3,4,5,6\}, \{1\}\}$

② $\Omega = \mathbb{R}$

The smallest sigma algebra that contains all intervals of the form (a, b) for $a, b \in \mathbb{R}$ is called the Borel sigma algebra.

https://en.wikipedia.org/wiki/Borel_set

Lebesgue

Probability measure : $P : \mathcal{F} \to [0,1]$

① $P(\Omega) = 1$

② $A_1, A_2, \ldots \in \mathcal{F}$ (countable collection of events)
disjoint

$$\Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Q: Why not assign probability measure on outcomes?
(Eg: Uniform distribution on $[0,1]$)

Q: Why not always choose $\mathcal{F}$ = power set of Ω?
A: Not possible to have a consistent measure
Eg: Vitali set https://en.wikipedia.org/wiki/Vitali_set

Random variable:

$$X : \Omega \to \mathbb{R}$$

s.t $X^{-1}((a,b]) \in \mathcal{F}$

(every interval corresponds to a valid event)

Example: Infinite sequence of coin tosses

H T H H H - —

0 . 0 1 0 0 0

Say $b_i = \begin{cases} 0 & \text{if the ith toss} = H \\ 1 & \text{if } T \end{cases}$

$0 \cdot b_1 b_2 b_3 \cdots$

$$x = b_1/2 + b_2/4 + b_3/8 + \cdots$$

$$= \sum_{i=1}^{\infty} b_i / 2^i$$

$$X = \sum_{i=1}^{\infty} b_i / 2^i$$

$$\Pr[X \in [0, 1/2)] = 1/2$$

$$\Pr[X \in [1/4, 1/2)] = 1/4$$

$$\Pr[X \in [a, b)] = b - a \longrightarrow \text{Uniform distribution on } [0, 1]$$

$$\Pr\Big[X \in \underbrace{\Big[j/2^i, \frac{j+1}{2^i}\Big)}_{b_1, b_2 \ldots b_i}\Big] = 1/2^i$$

$$[1/3, 2/3)$$

Probability measure :

For a random variable, specified by the cumulative distribution function (cdf)

$$F_X(x) = \Pr[X \leq x]$$

$(a, b]$

$(-\infty, b]$

Discrete : - Probability mass function

$$p_X(x) = \Pr[X = x] \rightarrow$$ Assign probabilities to outcomes

Continuous : Probability density function

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Common distributions : Make sure that you know the pmf/pdf :

1. Bernoulli
2. Binomial
3. Poisson
4. Uniform (discrete & continuous)
5. Gaussian
6. Laplace
7. Gamma
8. Chi-squared.

Expectation

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \, p_X(x) \quad , \text{ if discrete}$$

$$= \int_{-\infty}^{\infty} x \, f_X(x) \, dx \quad , \text{ if continuous}$$

$$\mathbb{E}[g(X)] = \begin{cases} \sum_{x \in \mathcal{X}} g(x) \, p_X(x) \\ \int_{-\infty}^{\infty} g(x) \, f_X(x) \, dx \end{cases}$$

- Mean : $\mathbb{E}X = \mu$
- Variance : $\mathbb{E}[(X - \mathbb{E}X)^2] = \sigma^2$

Moment generating function

k'th moment : $\mathbb{E}[X^k]$ → noncentral

$\mathbb{E}[(X-\mu)^k]$ → central.

MGF : $$\mathbb{E} e^{\lambda X} = \begin{cases} \sum_{x \in \mathcal{X}} e^{\lambda x} p_X(x) \\ \int_{-\infty}^{\infty} f_X(x) e^{\lambda x} \, dx \end{cases}$$

Exercise : Compute the MGF for all the common distributions listed above.

Basic Tail bounds: $\Pr[X \geq \delta]$

Markov inequality: Let X be a non-negative random variable (i.e., $\Pr[X<0]=0$) & $\mathbb{E}X = \mu < \infty$

Then,

useful only if $t \geq \mu$

$$\Pr[X \geq t] \leq \frac{\mu}{t} \quad \forall t > 0$$

Proof:

$$\mu = \int_0^\infty x f_X(x)\,dx = \underbrace{\int_0^t x f_X(x)\,dx}_{\geq 0} + \int_t^\infty x f_X(x)\,dx$$

$$\frac{\mu}{t} \geq \int_t^\infty \frac{x f_X(x)}{t}\,dx \geq \frac{t}{t}\int_t^\infty f_X(x)\,dx = \frac{\cancel{t}}{\cancel{t}}\Pr[X \geq t]$$

Chebyshev inequality:

Let X be a rv with $EX = \mu < \infty$,

$$\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right] < \infty.$$

Then,

$$\Pr\left[\,|X-\mu| > t\right] \le \frac{\sigma^2}{t^2} \qquad t > 0$$

Proof:

$$\Pr\left[|X-\mu| > t\right] = \Pr\Big[\underbrace{(X-\mu)^2}_{Y} > t^2\Big]$$

$$\le \frac{\mathbb{E}Y}{t^2} = \frac{\sigma^2}{t^2}.$$

Sequences of random variables

$X_1, X_2, X_3 \cdots X_n$ are independent & identically distributed if

$$p_{X_1 \cdots X_n}(x_1, \cdots, x_n) = \prod_{i=1}^{n} p_X(x_i)$$

$$f_{X_1 \cdots X_n}(x_1 \cdots x_n) = \prod_{i=1}^{n} f_X(x_i)$$

Q: Let $Y = \frac{1}{n} \sum_{i=1}^{n} X_i$ where $X_1 \cdots X_n$ are iid

What is $\mathbb{E}Y$? $= \mathbb{E}X = \mathbb{E}X_1$

What is $\mathrm{Var}(Y)$? $= \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) = \sigma^2/n$

$$\Pr\Big[\Big|\underbrace{\frac{1}{n} \sum_{i=1}^{n} X_i}_{Y_n} - \mu\Big| > \epsilon\Big] \leq \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty$$

Convergence of random variables

① We say that $X_1, X_2, X_3 \dots$ converges to X in probability if

$$\Pr[\,|X_n - X| > \epsilon\,] \to 0 \quad \text{as } n \to \infty \qquad \text{for all } \epsilon > 0$$

We denote this as $X_n \xrightarrow{P} X$

② We say that $X_1 X_2 \dots$ converges to X almost surely / with probability 1 if

$$\Pr[\lim_{n\to\infty} X_n = X] = 1 \iff \Pr[\lim_{n\to\infty} |X_n - X| = 0] = 1$$

We denote this as $X_n \xrightarrow{a.s.} X$

(3) We say that $X_1 X_2 X_3 \ldots$ converge to X in distribution if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$ at all points where F_X is continuous.

We denote this $X_n \xrightarrow{d} X$

(4) We say that $X_1 X_2 \ldots$ converges to X in L^p (for $p \geq 1$) if

$$\lim_{n\to\infty} E\left[|X_n - X|^p\right] = 0$$

Examples :

① $(X_n, n \geq 1)$ independent

$$\Pr[X_n = 0] = 1 - 1/n \qquad \Pr[X_n = 1] = \frac{1}{n}$$

$$\Pr[X = 0] = 1 \qquad F_X(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

① $$F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ 1 - 1/n, & x \in [0, 1) \\ 1 & x \geq 1 \end{cases}$$

$$\lim_{n \to \infty} F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} = F_X(x)$$

$$\therefore X_n \xrightarrow{d} X$$

② $\Pr[\,|X_n - X| > \epsilon\,] = \Pr[\,|X_n| > \varepsilon\,] = 1/n$

$\longrightarrow 0$

as $n \to \infty$

③ $\mathbb{E}|X_n - X|^p = \mathbb{E}|X_n|^p = 0^p(1 - 1/n) + 1^p \frac{1}{n}$

$= 1/n \longrightarrow 0$ as $n \to \infty$

④ $\Pr[\lim_{n\to\infty} |X_n - X| = 0]$ $\quad X_n \overset{a.s}{\not\to} 0$

$\Pr[\,X_n < \varepsilon \text{ for all } n \geq N\,] = \Pr[X_N = 0, X_{N+1} = 0, \ldots]$

$= (1 - 1/N)\left(1 - \frac{1}{N+1}\right)\left(1 - \frac{1}{N+2}\right) \cdots = 0$

$\Pr[X_n = 1] = e^{-n}$ $\Pr[X_n = 0] = 1 - e^{-n}$

$-$ $\Pr[X_n = 1/n] = 1/n$, $\Pr[X_n = 0] = 1 - \frac{1}{n}$

For given $\varepsilon > 0$, can we find N so

$$\Pr[X_n > \varepsilon \text{ for } n \geq N] = 0 \qquad \text{Yes.}$$

$$X_n \xrightarrow{as} X$$

(2) $X_1 \sim Ber(1/2)$

$$X_n = \begin{cases} X_1 & \text{if } n \text{ is odd} \\ \bar{X}_1 & \text{if } n \text{ is even} \end{cases}$$

$Pr[\,|X_n - X| > \varepsilon\,]$ does not converge as $n \to \infty$ for any X

$$F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ 1/2, & x \in [0,1) \\ 1, & x \geq 1 \end{cases}$$ for all n

Note :

$$\begin{array}{ccccc} & & X_n \xrightarrow{L^p} X & & \\ & & \Downarrow \;\; \not\Uparrow & & \\ X_n \xrightarrow{a.s} X & \begin{array}{c}\Longrightarrow \\ \not\Longleftarrow\end{array} & X_n \xrightarrow{P} X & \begin{array}{c}\Longrightarrow \\ \not\Longleftarrow\end{array} & X_n \xrightarrow{d} X \end{array}$$

in general.

Weak law of large numbers (WLLN)

If $X_1, X_2, \ldots$ are iid with mean μ, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

$$\Pr\left[\,\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right] \le \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty$$

Strong law of large numbers (SLLN)

If $X_1, X_2, \ldots$ are iid with mean μ, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s} \mu$$

for any $\varepsilon > 0$, $\delta > 0$, $\exists\, N$ st

$$\Pr\left[\,\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \le \varepsilon \text{ for all } n \ge N\right] \ge 1-\delta$$

Proof of WLLN if $Var(X_i) < \infty$

Central limit theorem

If $X_1, X_2 \ldots$ are iid with finite mean μ & finite variance σ^2,

then

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