Lecture 2

Aug 30, 2023

Concentration of random variables.

Questions: { Estimating average height of students
exit polls

n samples:

$$X_1, X_2, \dots, X_n \sim P$$

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow \mu := \mathbb{E}_{X \sim P} [X]$$

$\boxed{\text{Goal}}$ measure how much $\overline{X}_n$ deviates from μ

Law of Large numbers

(weak) $\forall \varepsilon \quad \lim_{n \to \infty} \Pr[|\overline{X}_n - \mu| < \varepsilon] = 1$

(strong) $\Pr[\lim_{n \to \infty} \overline{X}_n = \mu] = 1$

Central Limit Theorem:

$\mathrm{Var}_{X\sim P}[X]$

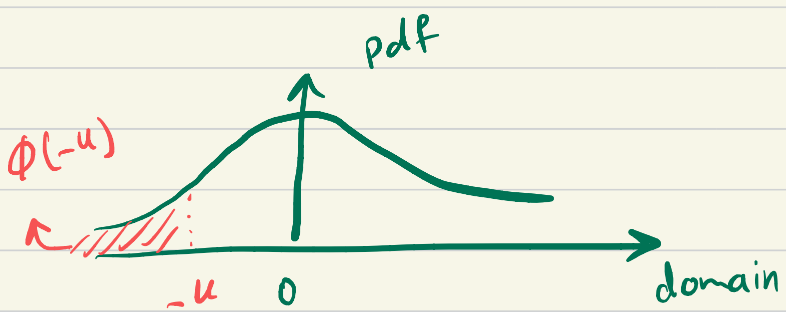
$$\sqrt{n}\,(\bar{X}_n - \mu) \to \mathcal{N}(0, \sigma^2)$$

$Z \sim \mathcal{N}(0,1)$

$$\Pr\left[\frac{\sqrt{n}\,|\bar{X}_n - \mu|}{\sigma} > u\right] \approx \Pr[|Z| > u]$$

$$= 2\Phi(-u)$$

where Φ is the cdf of the standard normal dist.



Look up table

$u = 1.96 \rightsquigarrow 2\Phi(-u) \simeq 95\%$

Hence: with prob. 0.95

$$\mu \in [\bar{X}_n - 1.96\,\sigma/\sqrt{n},\ \bar{X}_n + 1.96\,\sigma/\sqrt{n}]$$

[show plots]

- Quality of Approximation varies depending on P.

These are asymptotic results. Very general, but

- work in the limit,
- Do not indicate the relationship among the parameters:

n, d, ε, δ ?

d: dimension

ε: error

δ: confidence (in our example δ was $1-0.95 = 0.05$)

what about finite sample setting ?

Usefull tools to show concentration (tail bounds)

Markov's inequality:

For non-negative random variable X, and $a>0$:

$$\Pr[X \geq a] \leq \frac{E[X]}{a}$$

proof.

pdf

$$E[X] = \int_0^\infty x \Pr[X=x]\, dx$$

$$= \int_0^a x \Pr[X=x]\, dx + \int_a^\infty x \Pr[X=x]\, dx$$

$$\geq 0 + \int_a^\infty a \Pr[X=x]\, dx$$

$$\geq a \Pr[X \geq a]$$

$$\Rightarrow \Pr[X \geq a] \leq \frac{E[X]}{a} \qquad \square$$

Chebyshev's inequality

For a random variable with finite mean and variance, and $k > 0$:

$$\Pr[\,|X - \mathbb{E}[X]| \geq k\underbrace{\sigma}_{\text{standard deviation of } X}\,] \leq \frac{1}{k^2}$$

proof:

$$\Pr[\,|X - \mathbb{E}[X]| \geq k\sigma\,]$$

$$= \Pr[\,(X - \mathbb{E}[X])^2 \geq k^2\sigma^2\,]$$

$$\underbrace{\leq}_{\text{Markov}} \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2} \quad \square$$

Chernoff bound:

general structure of the proof:

For all $\varepsilon > 0$, $t > 0$:

$$\Pr[X \geq \varepsilon] = \Pr[e^{tX} \geq e^{t\varepsilon}]$$

$$\leq \frac{\mathbb{E}[e^{tX}]}{e^{t\varepsilon}} = e^{-t\varepsilon} M_X(t)$$

Markov

moment generating func

Since the bound holds for any t, we can conclude:

$$\Pr[X \geq \varepsilon] \leq \inf_{t>0} e^{-t\varepsilon} M_X(t) \quad \square$$

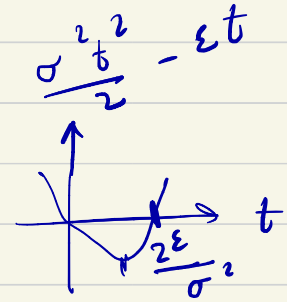
Example 1: standard normal

$$Z \sim N(1, 0)$$

$$M_Z(t) = E[e^{tZ}] = \exp\left(\frac{\sigma^2 t^2}{2}\right)$$

$$\Pr[Z > \varepsilon] \le e^{-t\varepsilon} M_Z(t)$$

$$= e^{-t\varepsilon} \exp\left(\frac{\sigma^2 t^2}{2}\right)$$



$$= \exp\left(\frac{\sigma^2 t^2}{2} - t\varepsilon\right)$$

$t := \varepsilon/\sigma^2$

$$= \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$

$$\Rightarrow \Pr[|Z - E[Z]| > \varepsilon]$$

$$\le 2\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$

Sub-Gaussian

The moment generating function determins concenteration.

What if X behave like a normal?

Definition A mean-zero random var. is sub-Gaussian with variance proxy s^2 if

$$M_X(t) \leq e^{s^2 t^2 / 2} \quad \forall t \in \mathbb{R}$$

[Hoeffding lemma]

$X \rightarrow$ zero mean random variable in $[a, b]$

$$M_X(t) := E[e^{tX}] \leq e^{t^2(b-a)^2/8}$$

$\Rightarrow$ Hence X is sub-Gassian where

$$s^2 = \frac{(b-a)^2}{4}$$

$$\longleftrightarrow$$

Suppose X_1 and X_2 are two independent sub-Gaussian random variables with variance proxies s_1^2 and s_2^2. Then

$X_1 + X_2 \rightarrow$ is sub-Gaussian with variance proxy $s_1^2 + s_2^2$

n independent random variables:

X_i with mean μ_i in $[a+\mu_i, b+\mu_i]$

$$\Pr\left[\frac{\sum_{i=1}^{n}(X_i-\mu_i)}{n} \geq \varepsilon\right]$$

$$\leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

Chernoff bound for Bernoulli Variables:

Suppose we have a coin with bias μ.
We flip this coin n times.
Let Y be the # heads we observed.
Then, we have:

$$\Pr\left[\frac{Y}{n} - \mu > \varepsilon \cdot \mu\right] \leq e^{-n\mu\varepsilon^2/3}$$

$$\Pr\left[\mu - \frac{Y}{n} < \varepsilon \mu\right] \leq e^{-n\mu\varepsilon^2/2}$$

Hoeffding bound:

$$\Pr\left[\frac{Y}{n} - \mu > \varepsilon\right] < e^{-2n\varepsilon^2}$$

$$\Pr\left[\mu - \frac{Y}{n} < \varepsilon\right] < e^{-2n\varepsilon^2}$$