Reminders

$X \sim N(\mu, \sigma^2) \quad Y = aX + b \quad Y \sim N(a\mu + b, a^2\sigma^2)$

$X_1, X_2, \ldots, X_n$ indep normal r.v.'s

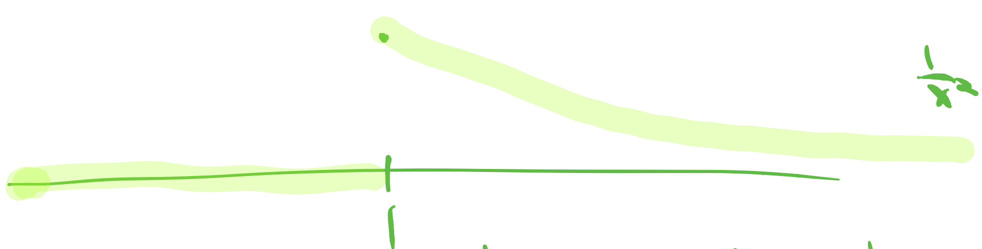
w/ $\mu_1, \mu_2, \ldots \mu_n \quad \sigma_1^2, \ldots \sigma_n^2$

$$Z = \sum_{i=1}^{n} X_i \qquad Z \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Example: $F_X(x) = \begin{cases} 1 - \frac{1}{x} & x \geq 1 \\ 0 & \text{o.w.} \end{cases}$

$$f_X(x) = \begin{cases} \frac{1}{x^2} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

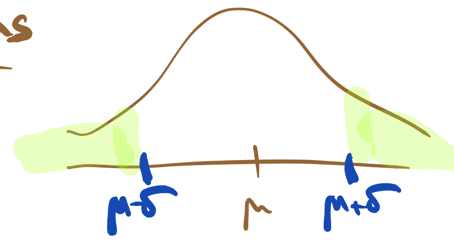
$$E(X) = \int_1^\infty x f_X(x)\,dx = \int_1^\infty \frac{x}{x^2}\,dx = \ln x \Big|_1^\infty$$



heavy-tailed dist'ns

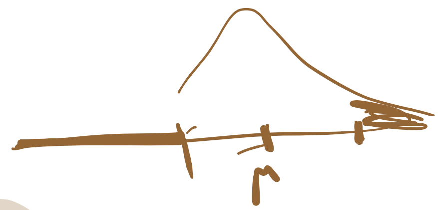
in what follows, I will assume finite mean & variance

Tail bounds and limit theorems



Markov's Inequality $\quad X \geq 0$

$$\Pr\left(X \geq \alpha E(X)\right) \leq \frac{1}{\alpha}$$



Chebychev's Inequality

Y mean μ, var σ^2

$$\boxed{\Pr\left(|Y - E(Y)| \geq \delta\right) \leq \frac{\sigma^2}{\delta^2}}$$

by Markov's inequality

$$\Pr\left(\left(Y-E(Y)\right)^2 \geq \alpha \underbrace{E\left[\left(Y-E(Y)\right)^2\right]}_{\sigma^2}\right) \leq \frac{1}{\alpha}$$

$$\parallel$$

$$\Pr\left(|Y-E(Y)| \geq \underbrace{\boxed{\sqrt{\alpha}\,\sigma}}_{\delta}\right) \leq \frac{1}{\alpha}$$

$$\sqrt{\alpha}\,\sigma = \delta \quad \Rightarrow \alpha = \frac{\delta^2}{\sigma^2}$$

$$\frac{1}{\alpha} = \frac{\sigma^2}{\delta^2}$$

Law of Large Numbers

$X_1, \ldots, X_n$ are $\underbrace{\text{independent, identically distributed}}_{\text{iid}}$

with $E(X_i) = \mu$
$\quad\;\; Var(X_i) = \sigma^2$

then $\displaystyle \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{} \mu$

$$\text{mean } \mu, \quad \text{variance } \frac{\sigma^2}{n}$$

$$\mathrm{Var}(\Sigma X_i) = n\sigma^2$$

linearity of var for indep r.v.'s.

$$\mathrm{Var}\left(\frac{1}{n}\Sigma X_i\right) = \frac{1}{n^2}\mathrm{Var}(\Sigma X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\Pr\left(\left|\frac{1}{n}\Sigma X_i - \mu\right| \geq \alpha\right) \leq \frac{\mathrm{Var}\left(\frac{1}{n}\Sigma X_i\right)}{\alpha^2} = \frac{\sigma^2}{n\alpha^2} \xrightarrow[n\to\infty]{} 0$$

Chebyshev's Inequality

Central Limit Theorem

$X_1, \ldots, X_n$ iid $E(X_i) = \mu$ $\mathrm{Var}(X_i) = \sigma^2$

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow[n\to\infty]{} N(0,1)$$

$$\Pr\left(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\,\sigma} \leq a\right) \xrightarrow[n\to\infty]{} \underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}\,dx}_{\Phi(a)}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Polling

Population of size N

p approve of Trump

$1-p$ don't

Want to determine p by polling

call n people (each indep., random draw from population)

$$X_i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ person approves} \\ 0 & \text{o.w.} \end{cases}$$

$E(X_i) = p$

$Var(X_i) = p(1-p)$

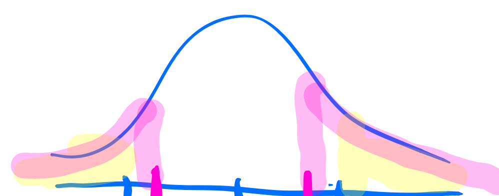
$\frac{1}{n}\sum_{i=1}^{n} X_i$ fraction of people we call that approve.

$$\sim N\left(p, \frac{p(1-p)}{n}\right)$$

$\hat{X}$: sample mean.

$$\Pr\left(|\hat{X}-p| > \underbrace{0.05}_{\varepsilon}\right)$$

$$= \Pr\left(\left|\frac{\hat{X}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| > \frac{0.05}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

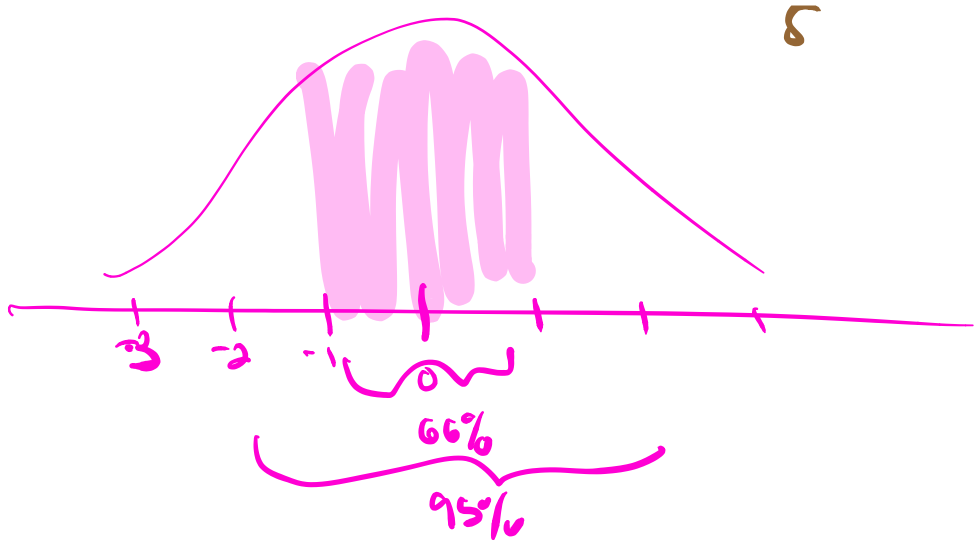


$$= \Pr\left(|\underset{N(0,1)}{Z}| > \frac{0.05}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$\frac{0.05}{\sqrt{\frac{1/4}{n}}}$$

$$\leq \Pr\left(|Z| > 0.1\sqrt{n}\right) \leq \underbrace{0.05}$$

5



if

$$0.1\sqrt{n} = 2$$

$$\sqrt{n} = 20$$

$$\boxed{n = 400}$$