18.650 : Fundamentals of Statistics

Statistics = analyze, describe, & reason about data

Types of Analysis:	Goal (Conclusion)
Descriptive Stats (mean, std, plots)	Summarize
Estimation	Say smthing about gen. population.
Confidence Intervals	Quantify uncertainty
Hypotheses Testing	Answer yes/no questions
Regression/Classification	Prediction
Causal Inference	Does x cause y?
Survival Analysis	predict "time to event"
Data Visualization (PCA, t-SNE, ...)	understand

i.i.d from $\mathbb{P}$

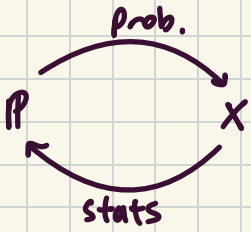
↳ independent & identically distributed

$X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathbb{P}$

X_i = effect of the drug on patient:

Pipeline: $X_i \rightarrow$ [Stats method] $\rightarrow \hat{\mathbb{P}} \approx \mathbb{P}$

Probability



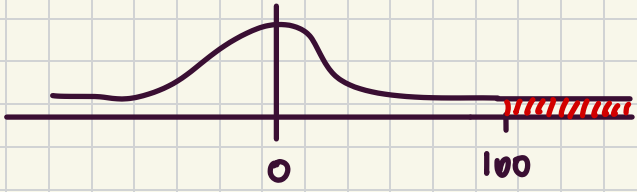
Example: $\mathbb{P} = N(0,1)$ $\quad X \sim \mathbb{P}$

prob: $\mathbb{P}(X \in [-3,3]) = ?$

stats: $X = 100$ Is it "likely" that $X \sim N(0,1)$?

↳ also need to def. "likely" (confidence)

$\mathbb{P}(X = 100) = 0$
$\mathbb{P}(X = 0) = 0$
} specific vals for continuous dist = 0



at least 100

$\mathbb{P}(X \geq 100) = 0.00001$ (very small...)

$\mathbb{P}(|X| \geq 100) = \ldots$

$X_1, \ldots, X_n \overset{iid}{\sim} P$

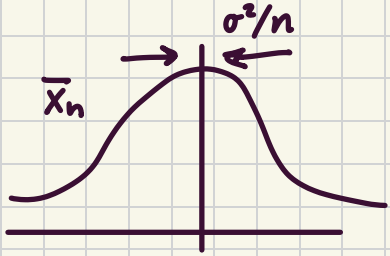
$\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$ (avg)

$E[X_1] = \mu$

$Var[X_1] = \sigma^2$

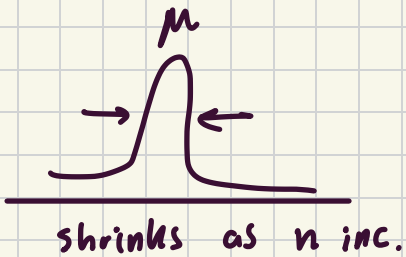
$E[\bar{X}_n] = \mu$

$Var[\bar{X}_n] = V[\bar{X}_n] = \frac{\sigma^2}{n}$



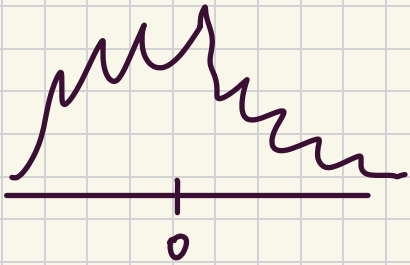
Law of Large #'s:

$\bar{X}_n \xrightarrow{\text{"}n \to \infty\text{"}} \mu$



$\bar{X}_n - \mu$ = new r.v w/ $\mu_i = 0$

$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ } std of $\bar{X}_n$



Central Limit Thm:

$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{"}n \to \infty\text{"}} N(0,1)$