Parameter Estimation

(+ a Mean Example)

Estimators

p Unknown distribution or population

θ Parameter of p wish to estimate Mean example μ σ, max, mode

Sample $X^n \stackrel{\text{def}}{=} X_1, X_2, \ldots, X_n \sim p$ ⫫ $X^3 = 5, -2, 6$

∧ for estimate

Estimator for θ function $\hat{\theta} : \mathbb{R}^n \to \mathbb{R}$ Maps X^n to $\mathbb{R}$

Upon observing X^n, estimate θ as $\hat{\theta}(X^n) \stackrel{\text{def}}{=} \hat{\Theta}$

μ	$\hat{\theta}(X^n)$:	$\frac{X_1+\ldots+X_n}{n}$	$\frac{\min\{X_i\}+\max\{X_i\}}{2}$	$X_1 \cdot X_2$
5, -2, 6	$\hat{\Theta}$:	3	2	-10

Observations

Distribution parameter θ Constant Mean 3.2

Estimate $\hat{\Theta} \stackrel{\text{def}}{=} \hat{\theta}(X^n)$ Random variable Ideally close to θ

Once sample X^n drawn Determines $\hat{\Theta}$

Single value 3.5 Point estimate vs. interval [3,4]

Any function is an estimator

How to

Come up with an estimator?

Evaluate its performance?


SPACEX
Elon Musk
Founder of SpaceX


SAMPLE X
Apply sample to any parameter X

Sample X

Property				
X			sample X	
min	$X_{\min}$	$\min_x \{x : p(x) > 0\}$	sample min	$\min_i \{X_i\}$
max	$X_{\max}$	$\max_x \{x : p(x) > 0\}$	sample max	$\max_i \{X_i\}$
mean	μ	$\sum_x x \cdot p(x)$	sample mean	$\frac{1}{n} \sum_{i=1}^{n} X_i$

Simple

If sample is whole population, exact

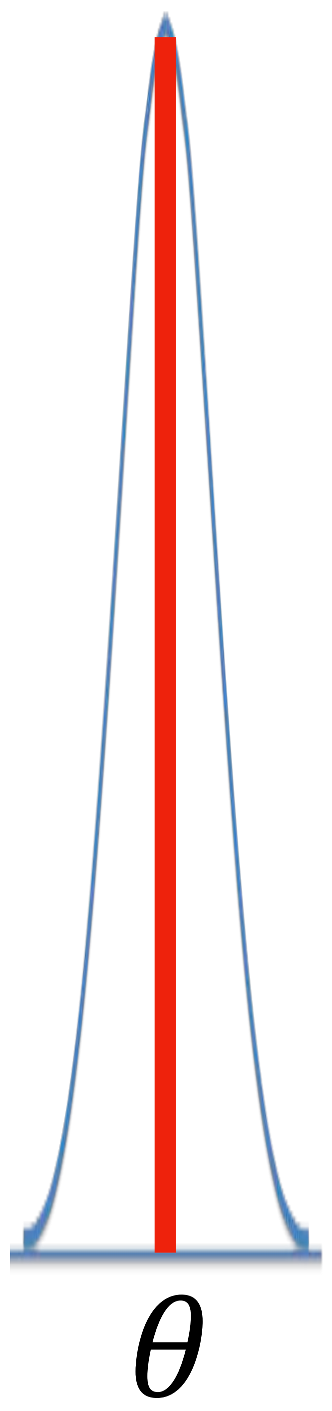
Sometimes works well

Even for small samples

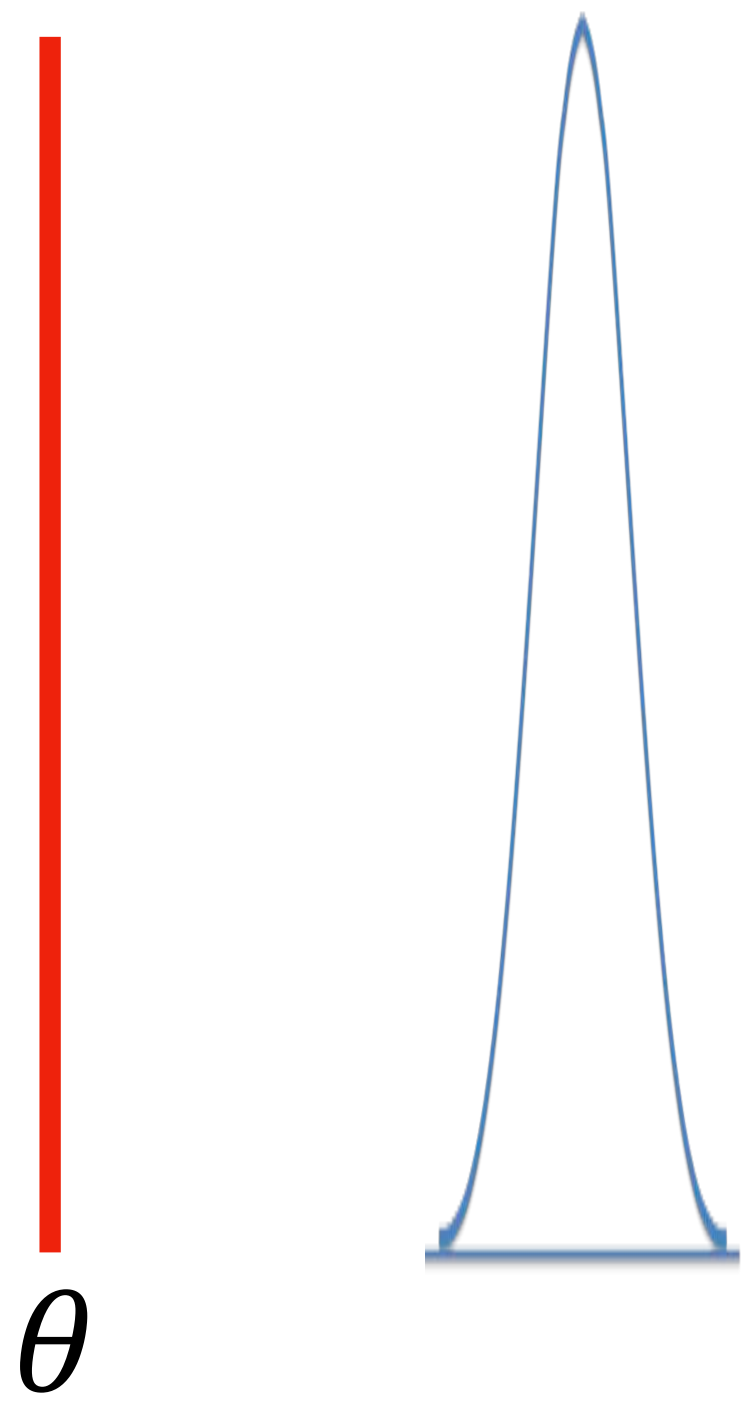
Estimator Evaluation

Parameter may have several estimators

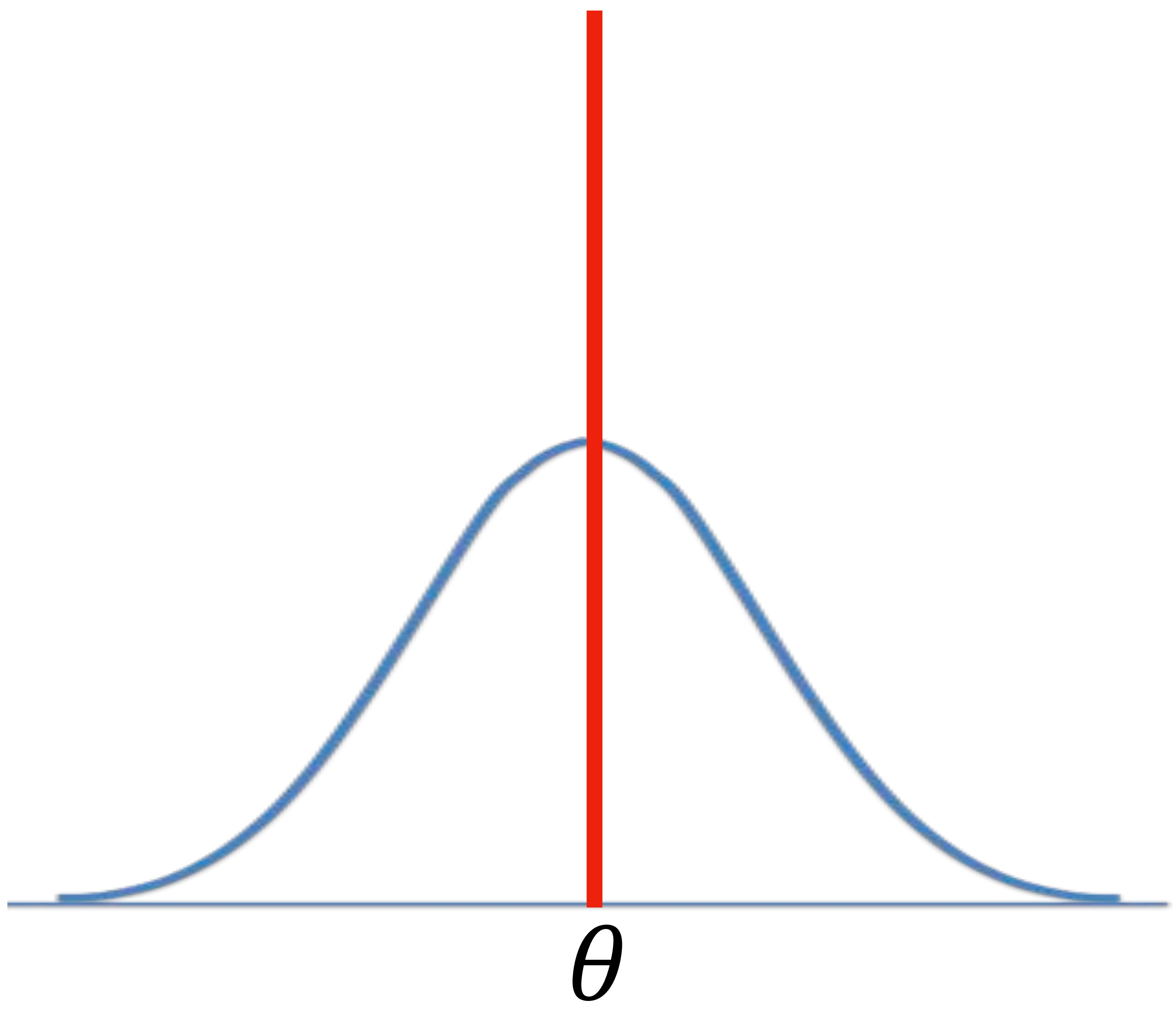
Evaluate quality of estimator for a parameter



Good



Bias



Variance

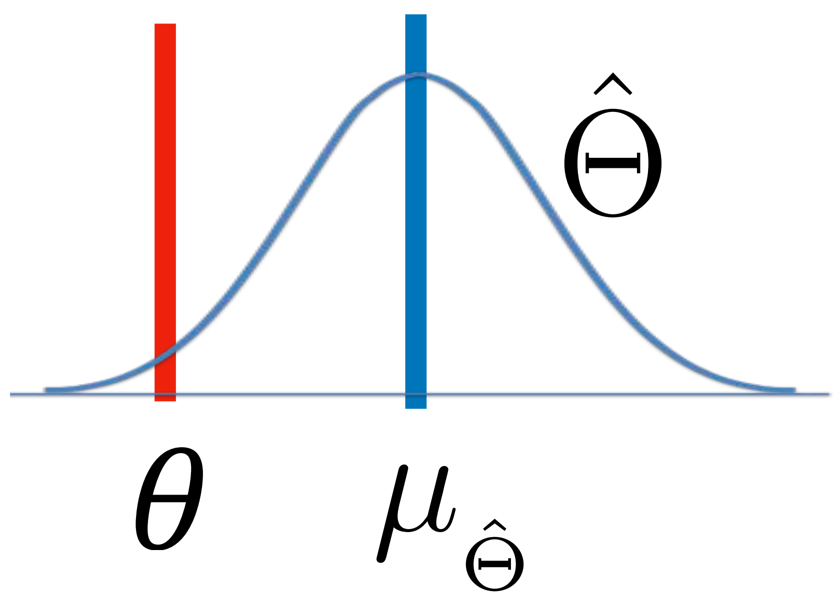
Bias

$\hat{\Theta}$ estimator for θ

Bias of $\hat{\Theta}$ is its expected overestimate of θ

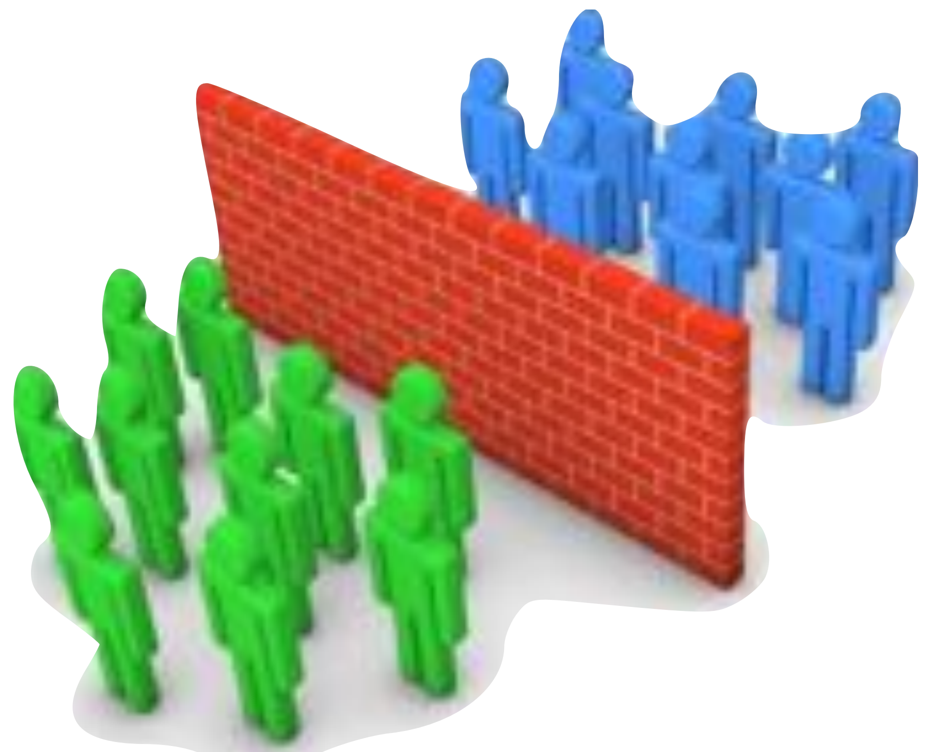
$$\mathrm{Bias}_\theta(\hat{\Theta}) \stackrel{\text{def}}{=} E(\hat{\Theta} - \theta) = \mu_{\hat{\Theta}} - \theta$$

$$\rightarrow \mathrm{Bias}(\hat{\Theta})$$



Estimator with 0 bias is unbiased

$$\mu_{\hat{\Theta}} = \theta$$



Bias = Inequality

Variance

$$V(\hat{\Theta}) = E(\hat{\Theta} - \mu_{\hat{\Theta}})^2$$

Unrelated to θ

Ideally 0 bias variance

Typically tradeoff

Mean Example

Unknown distribution or population p

Estimate mean μ

n samples

$$X_1, \ldots, X_n \sim p \;\perp\!\!\!\perp$$

Sample mean

$$\overline{X} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} X_i$$

Evaluate

Bias

Variance

Weak Law of Large Numbers

Sample Mean - Bias

Sample mean

$$\overline{X} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} X_i$$

Expectation

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$$

Bias

$$\mathrm{Bias}(\overline{X}) = E(\overline{X}) - \mu = \mu - \mu = 0$$

Sample mean is unbiased estimator for distribution mean

Sample Mean - Variance

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right)$$

$$\overset{\perp\!\!\!\perp}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2$$

$$= \frac{\sigma^2}{n}$$

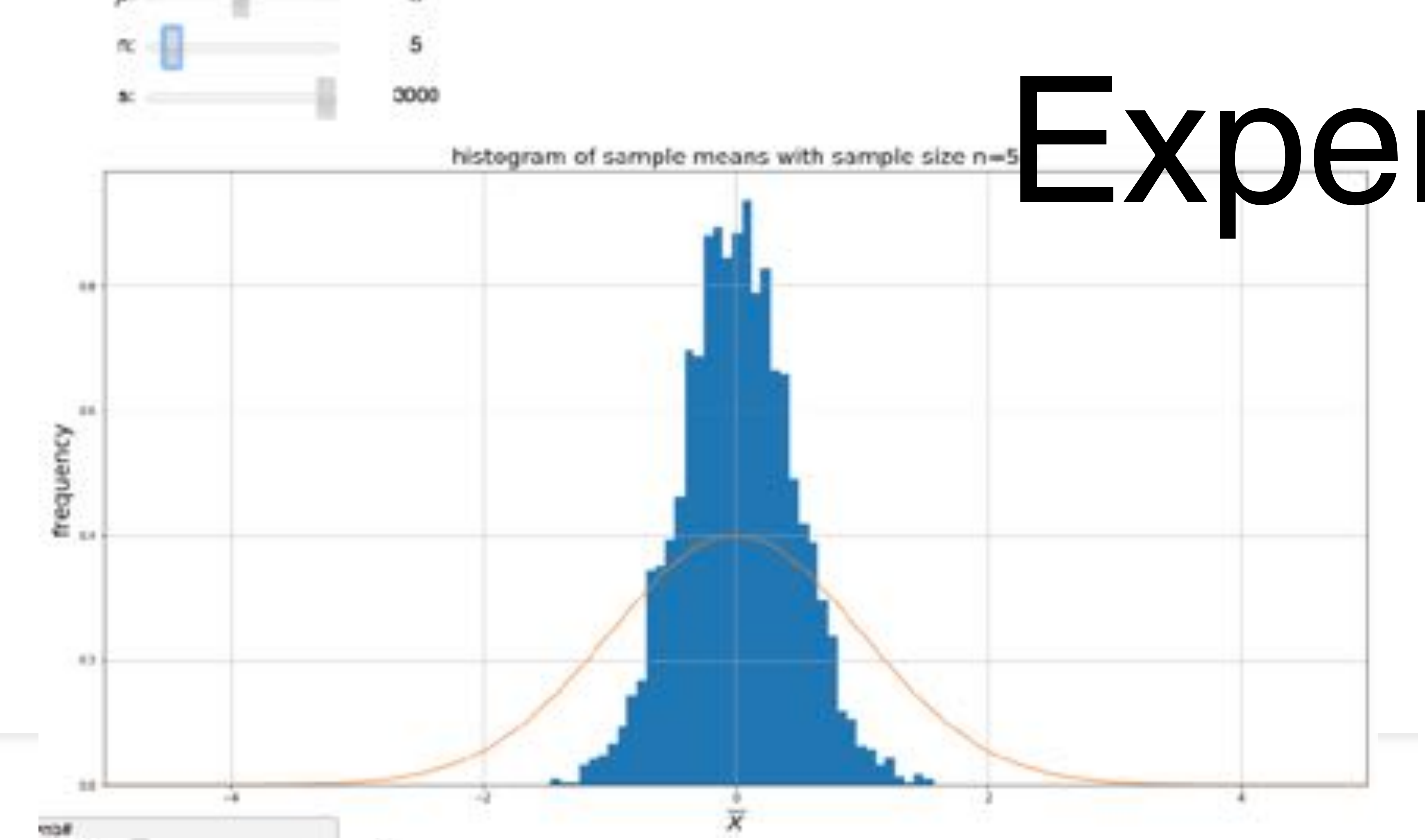
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

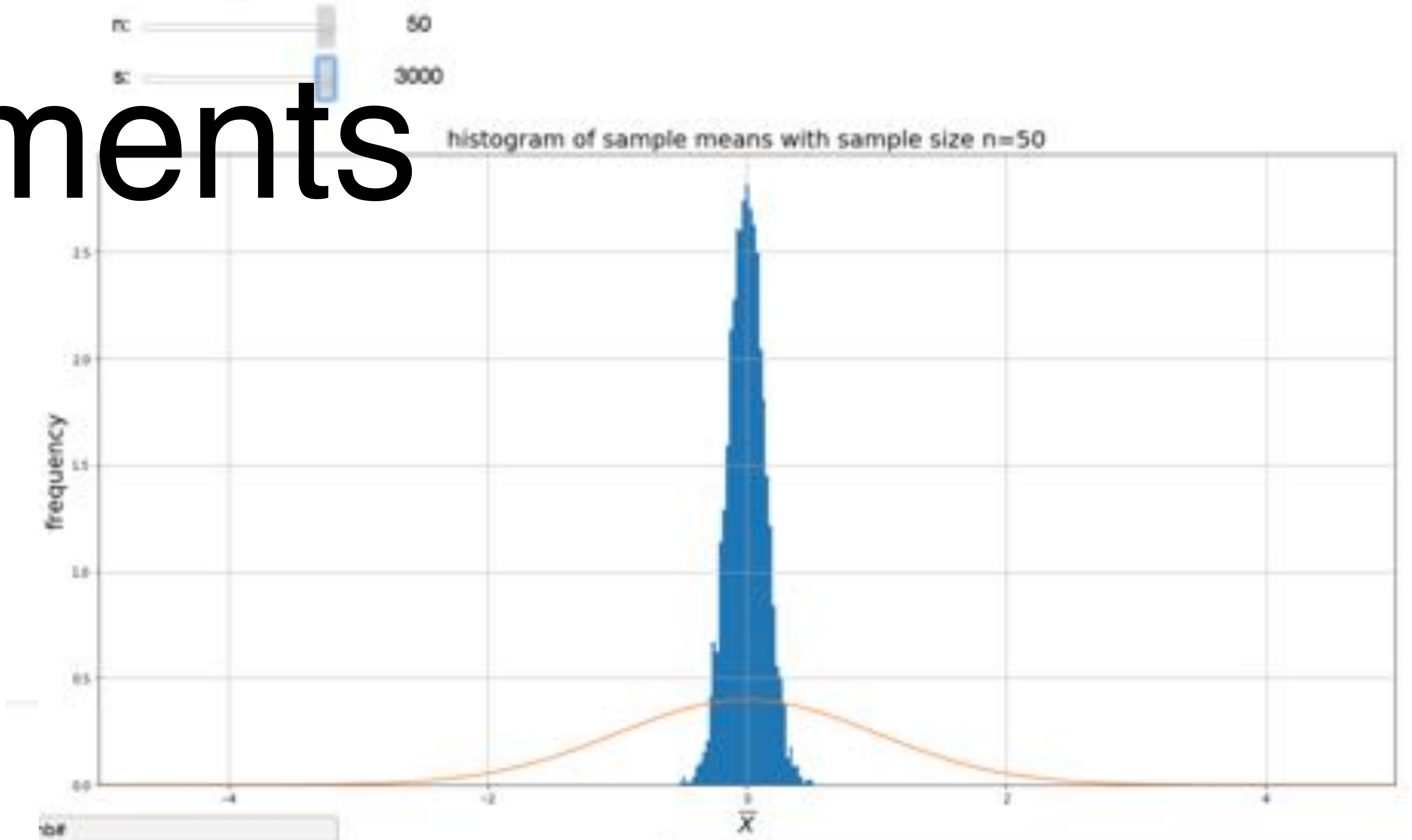
Will return

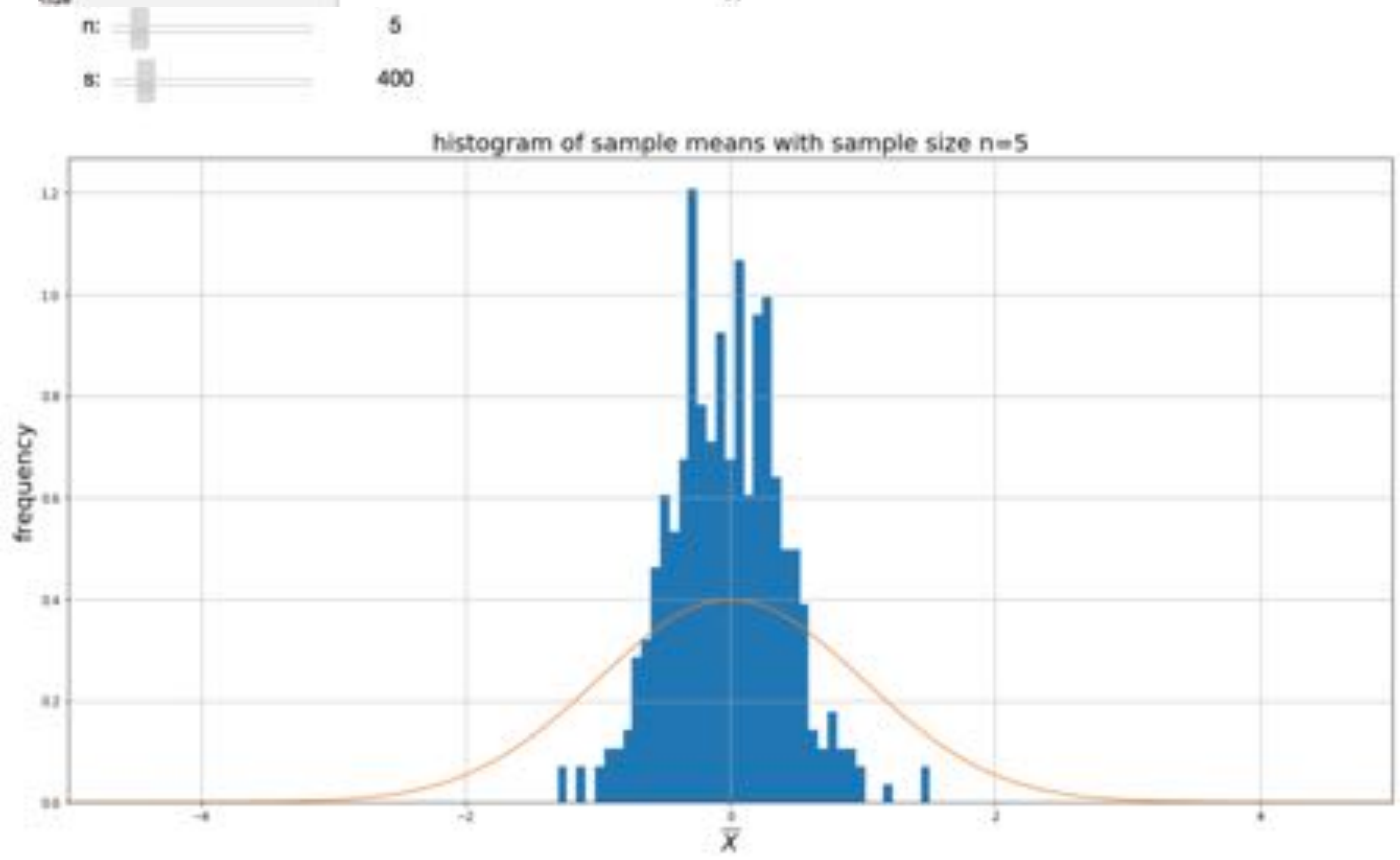
Increases with σ

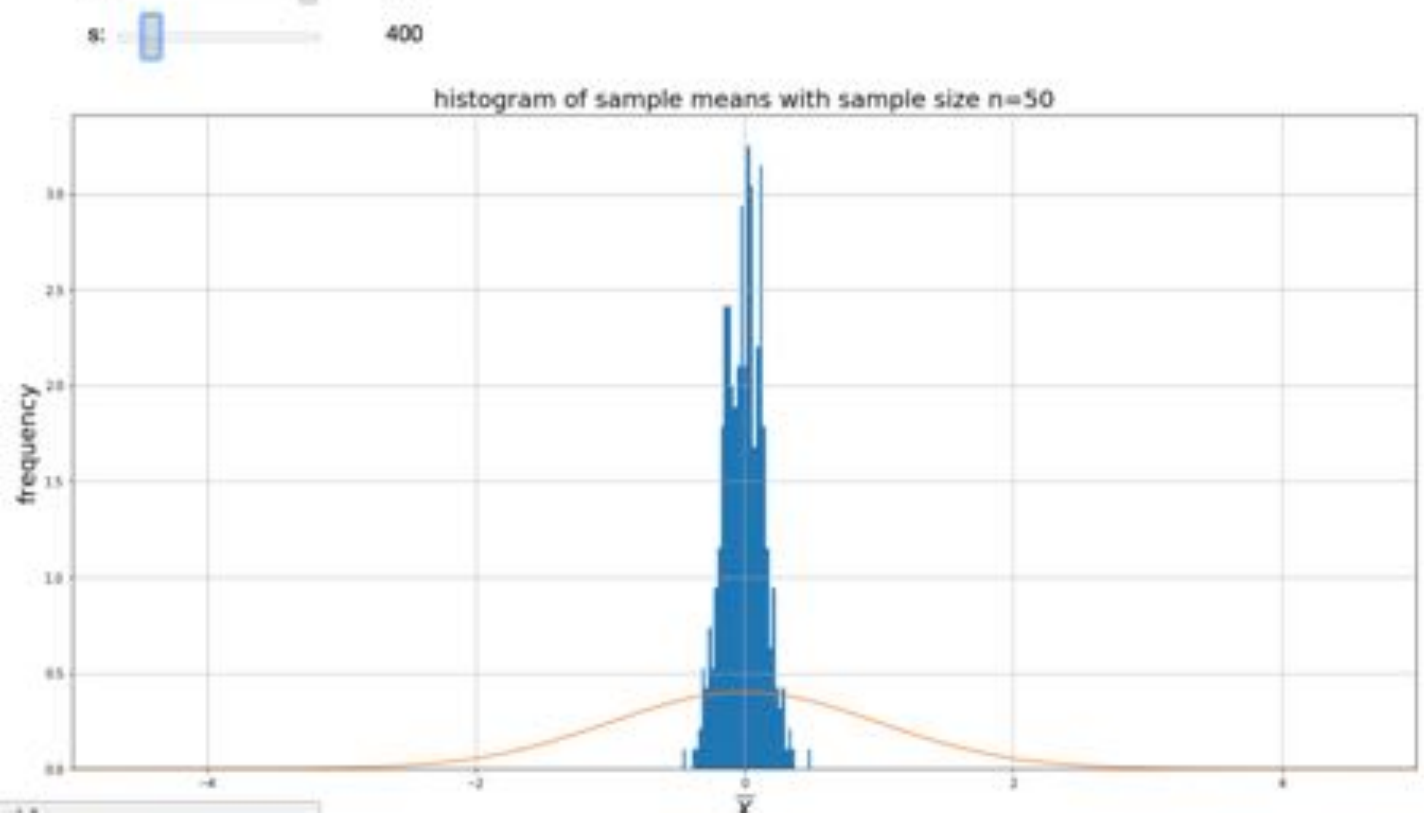
Decreases with n

Experiments









Mean Squared Error

Single measure for performance of estimator $\hat{\Theta}$ for θ

MSE of $\hat{\Theta}$ is its expected squared distance from θ

$$\mathrm{MSE}_\theta(\hat{\Theta}) \stackrel{\text{def}}{=} E(\hat{\Theta} - \theta)^2 \qquad \to \mathrm{MSE}(\hat{\Theta})$$

Common in science and engineering

Communication

Transportation

Production

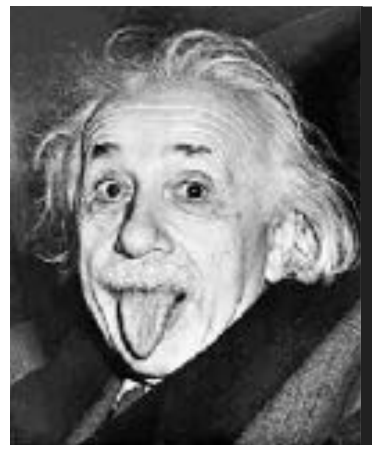
Need to re-evaluate?

Relate to bias and variance

Bias-Variance Bromance

MSE = Bias² + Variance

$$E(X^2) = E^2(X) + V(X)$$



$$\text{Energy} = \mu^2 + \sigma^2$$

$$\mathrm{MSE}(\Theta) = E(\Theta - \theta)^2$$

$$= E^2(\Theta - \theta) + V(\Theta - \theta)$$

$$E(\Theta - \theta) \stackrel{\text{def}}{=} \mathrm{Bias}(\Theta)$$

$$V(\Theta - \theta) = V(\Theta)$$

$$= \mathrm{Bias}^2(\Theta) + V(\Theta)$$

MSE of Sample Mean

$$\mathrm{MSE}_\mu(\overline{X}) \;=\; \mathrm{Bias}^2_\mu(\overline{X}) + V(\overline{X}) \;=\; \frac{\sigma^2}{n}$$

Increases with σ

Decreases with n

Same estimator works for all distributions

Accuracy (MSE) independent of population size

Parameter Estimation

(+ a Mean Example)