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

$\wedge$ 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?

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

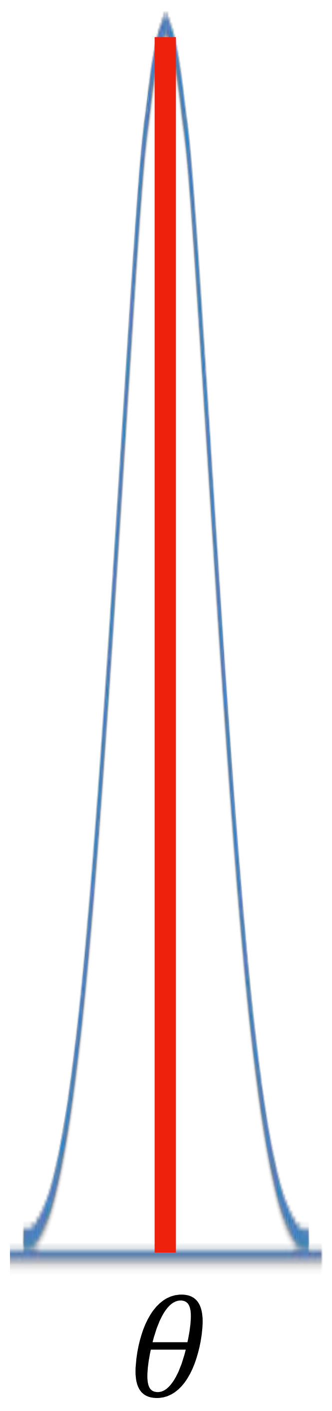
Sometimes works well

Even for small samples

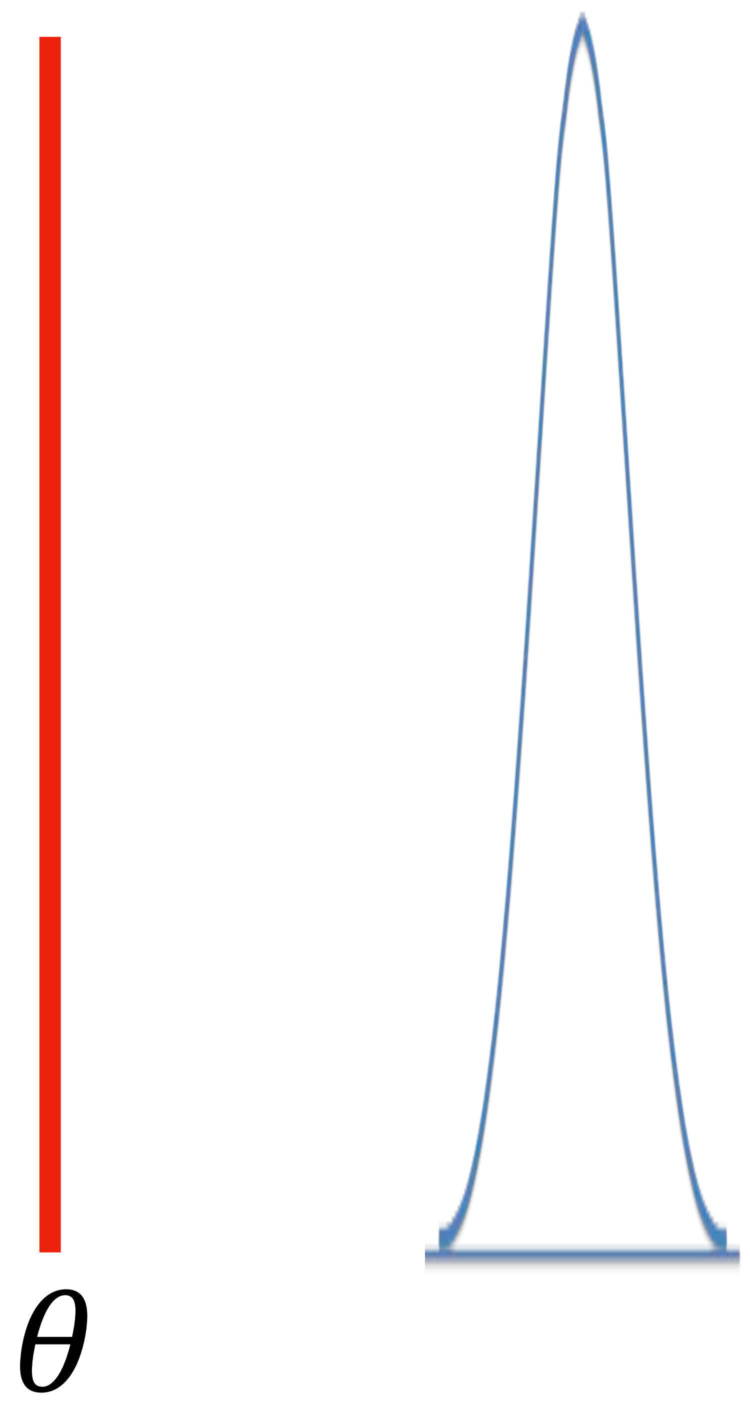
Estimator Evaluation

Parameter may have several estimators

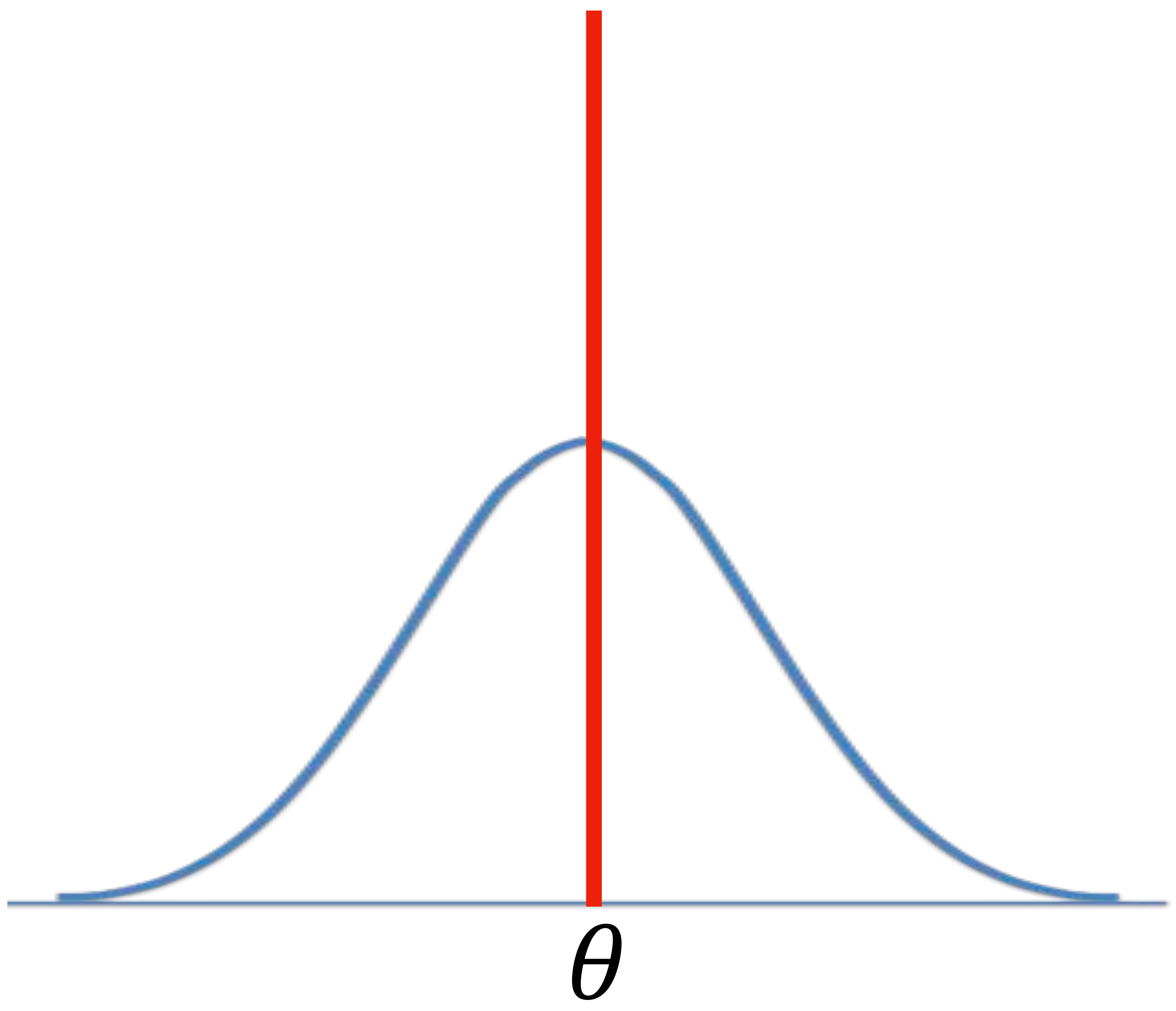
Evaluate quality of estimator for a parameter



Good



Bias



Variance

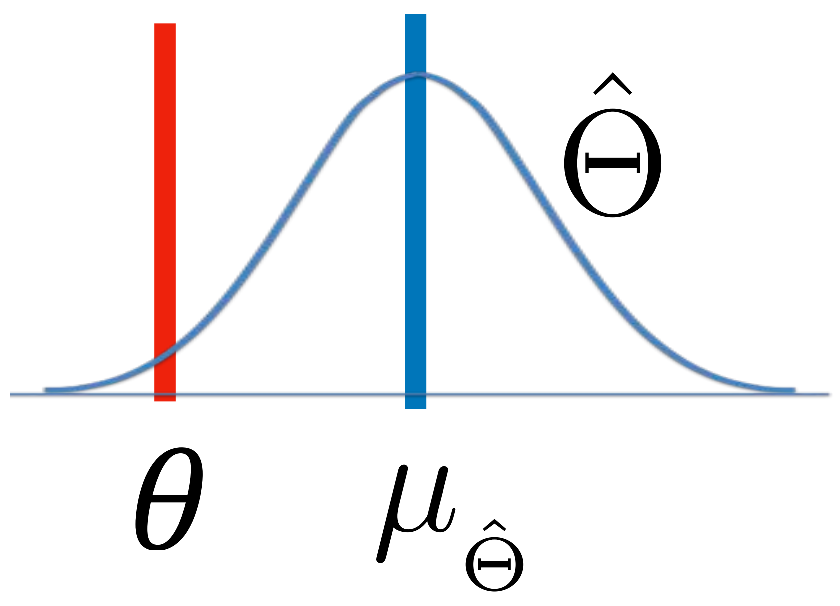
Bias

$\hat{\Theta}$ estimator for θ

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

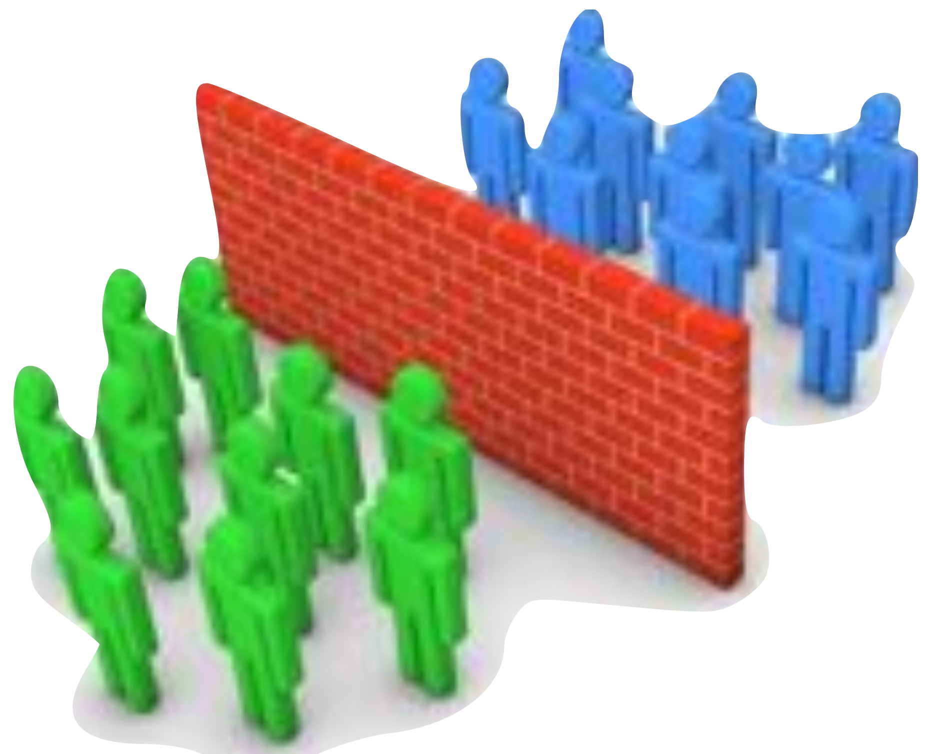
$$\text{Bias}_\theta(\hat{\Theta}) \stackrel{\text{def}}{=} E(\hat{\Theta} - \theta) = \mu_{\hat{\Theta}} - \theta$$

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



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

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

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

Sample mean is unbiased estimator for distribution mean

Sample Mean - Variance

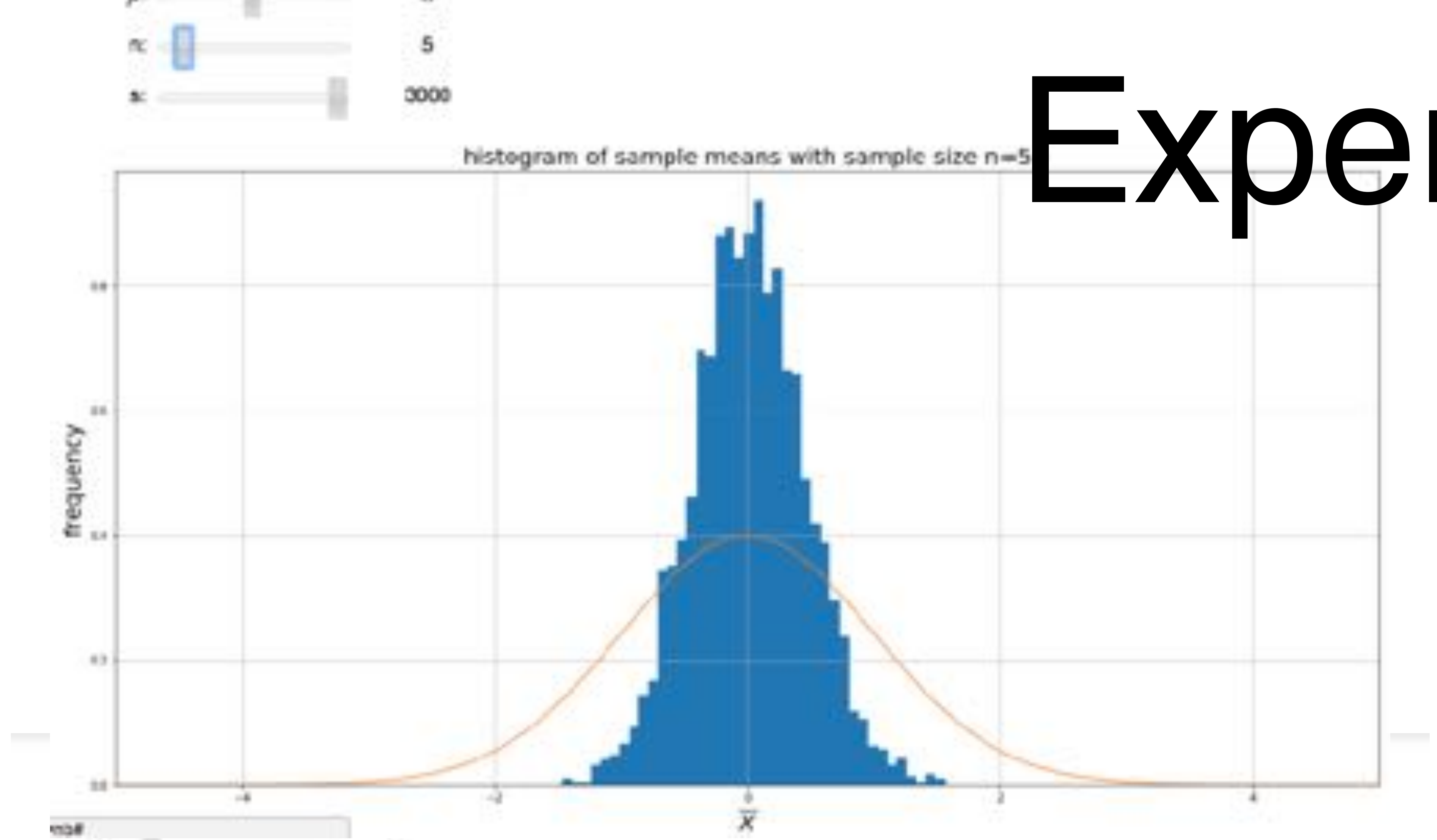
$$
\begin{aligned}
V(\overline{X}) &= V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) \\
&\overset{\perp\!\!\!\perp}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

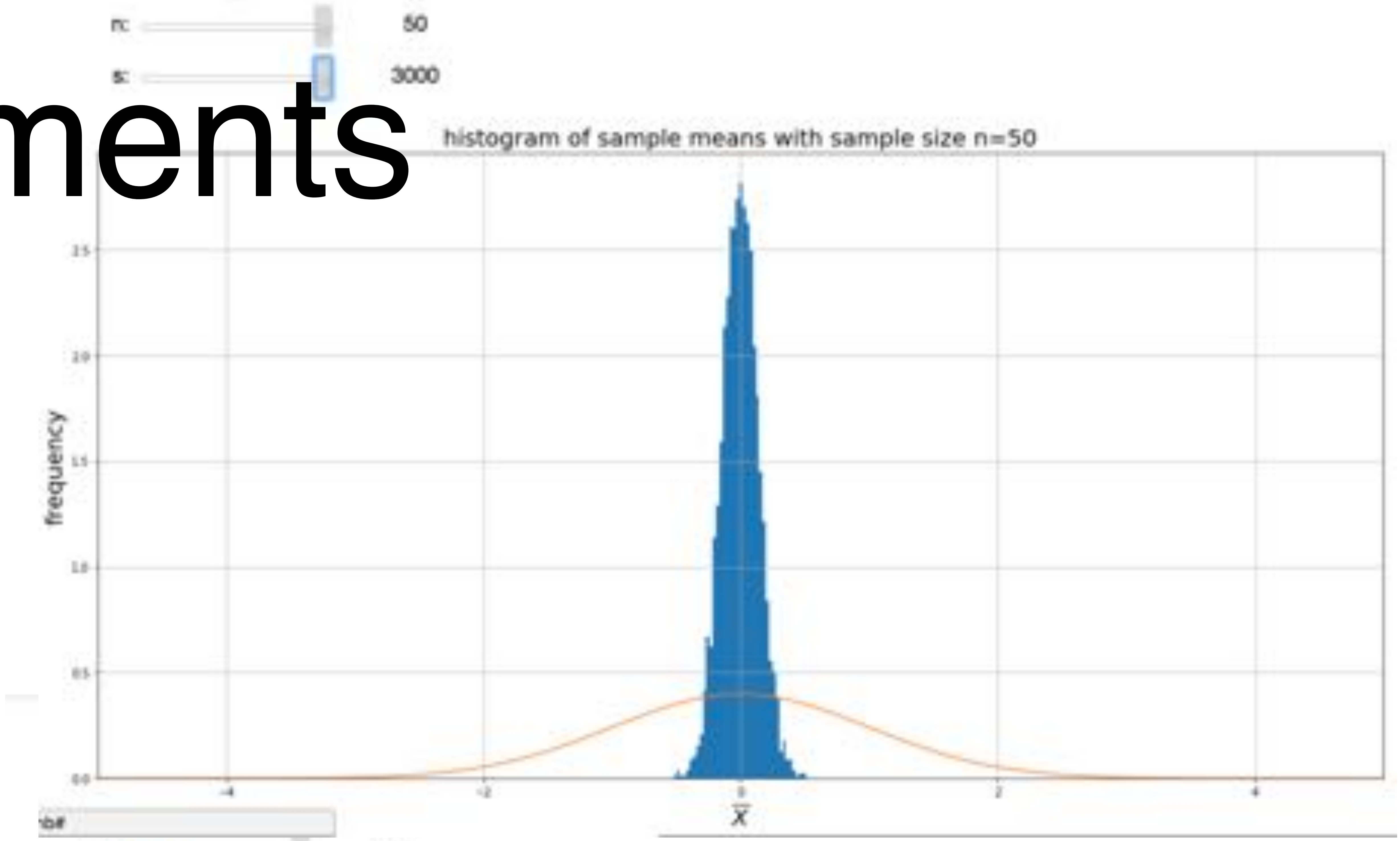
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

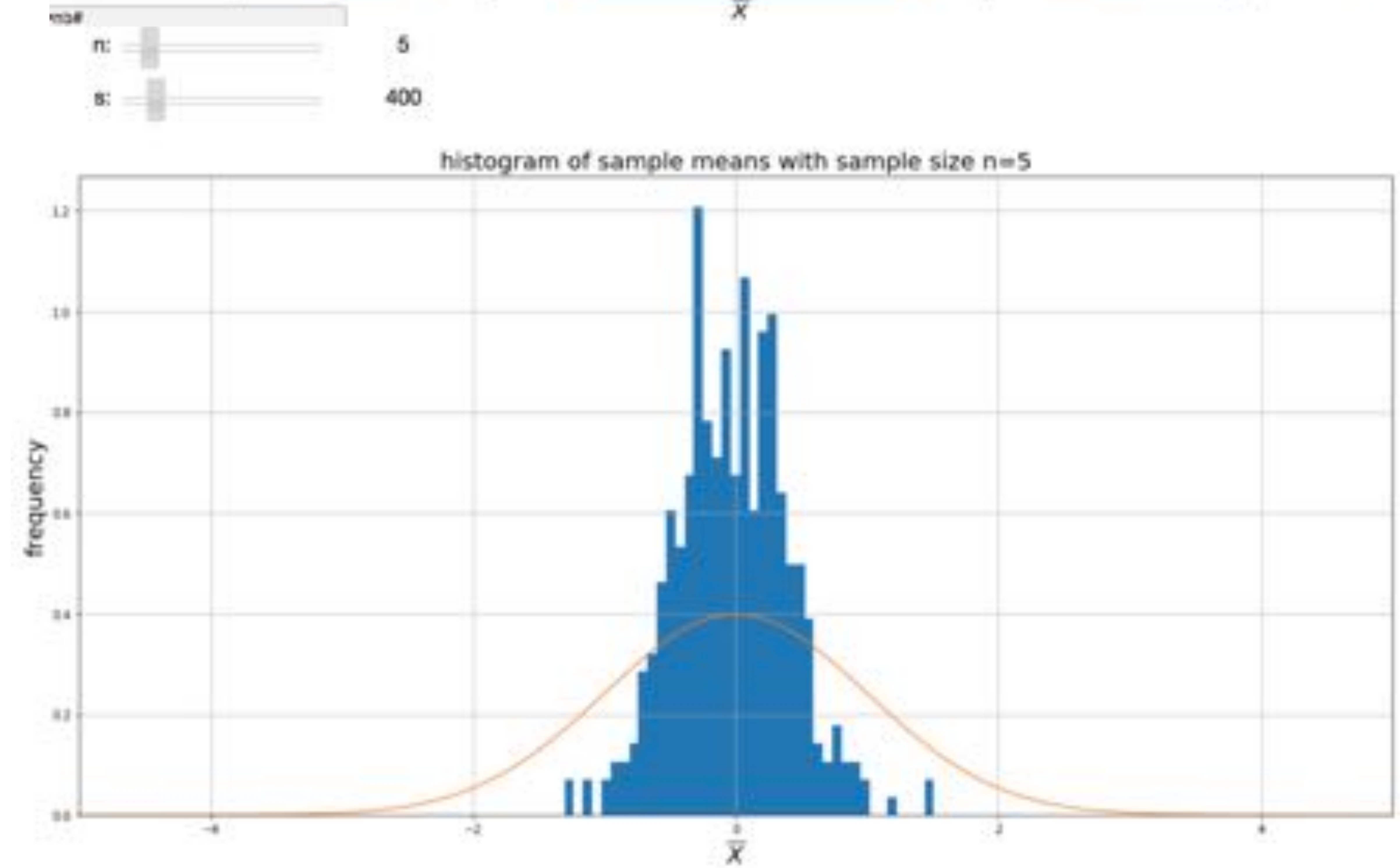
Increases with σ

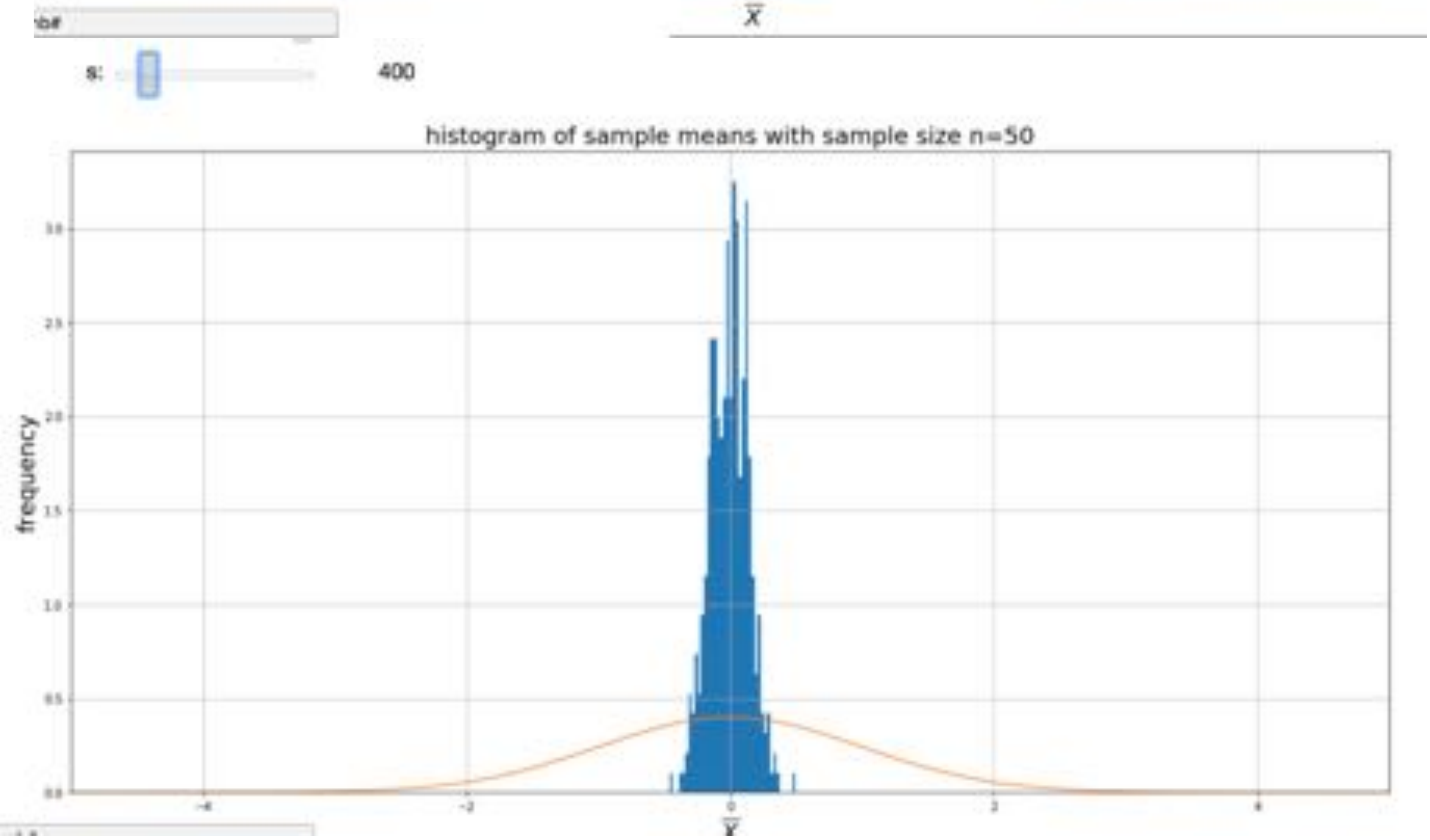
Decreases with n

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

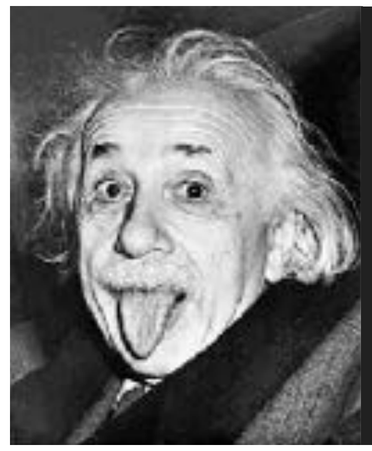
Communication Transportation Production

Need to re-evaluate? Relate to bias and variance

Bias-Variance Bromance

MSE = Bias² + Variance

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



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

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

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

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

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

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

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)

Variance Estimation

(It’s Almost Elementary)

Natural estimator

Two calculations

Biased?

Mystery

Estimating the Variance

Unknown distribution or population p

mean μ

variance σ^2

Estimate σ^2

Sample of n observations $X_1, \ldots, X_n \sim p \perp\!\!\!\perp$

No distribution

Expectation → average

Mean $\mu = E(X_i) \longrightarrow \overline{X} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} X_i$

Variance $\sigma^2 = E(X_i - \mu)^2 \longrightarrow \text{“}S^2\text{”} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

“Raw” sample variance

“Raw”

Soon

Random variable

ExSample

Sample

n=5 observations

2, 1, 4, 2, 6

$$\overline{X} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{2+1+4+2+6}{5} = \frac{15}{5} = 3$$

Estimated mean

$$\text{``}S^2\text{''} = \frac{1}{5}\sum_{i=1}^{5} (x_i - \overline{x})^2 = \frac{1+4+1+1+9}{5} = \frac{16}{5} = 3.2$$

Estimated variance

$V(X)$ = $E(X-\mu)^2$ = $E(X^2)-\mu^2$

Similar expression for "S^2"

One-Pass Calculation

$$\sum_{i=1}^{n} \to \sum$$

$$\begin{aligned}
\sum(x_i - \overline{x})^2 &= \sum\left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right) \\
&= \sum x_i^2 - 2\overline{x}\sum x_i + \sum \overline{x}^2 \\
&= \sum x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 \\
&= \sum x_i^2 - n\overline{x}^2
\end{aligned}$$

$$\sum x_i = n\overline{x}$$

$$\text{“}S^2\text{”} \stackrel{\text{def}}{=} \frac{1}{n}\sum(x_i - \overline{x})^2$$

Intuitive
Arguments

$$= \frac{1}{n}\sum x_i^2 - \overline{x}^2$$

Fewer subtractions
One pass

ExSample 

n=5

2, 1, 4, 2, 6

Saw

$$\overline{X} = 3$$

$$``S^2" = 3.2$$

One pass

$$\begin{aligned} ``S^2" &= \frac{1}{5}\sum_{i=1}^{5} x_i^2 - \overline{x}^2 \\ &= \frac{4+1+16+4+36}{5} - 3^2 \\ &= 12.2 - 9 \\ &= 3.2 \end{aligned}$$ ✅

Even more interesting…

“S²” Biased?

Mean

E → average of samples

Unbiased

Last lecture
WLLN

$$E(X_i) = \mu \qquad \overline{X} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} X_i \qquad E(\overline{X}) = \mu$$

Stands to reason

Variance

E → average of samples

Unbiased

?

$$E(X_i - \mu)^2 = \sigma^2 \qquad \text{“}S^2\text{”} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad E(\text{“}S^2\text{”}) = \sigma^2$$

But is it?

Sherlock Holmes, A study in Scarlet

"**No data yet**," he answered.
"It is a capital mistake to theorize before you have all the evidence. It **biases** the judgment."

Is "S^2" Biased?

Magnifying glass

→

Simulation

Simulation Plan

Pick a distribution — σ^2 known

Generate n observations — $X_1, \ldots, X_n$

Calculate $\text{“}S^2\text{”} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ — Check $E(\text{“}S^2\text{”}) \stackrel{?}{=} \sigma^2$

Find Expectation? — Last lecture, WLLN — E ≈ Average of many

r experiments — $\text{“}S^2\text{”}$ for each — Average $\overline{\text{“}S^2\text{”}}$ → $E(\text{“}S^2\text{”})$

Compare — Calculated $\overline{\text{“}S^2\text{”}}$ estimating $E(\text{“}S^2\text{”})$ — Known σ^2

Similar → Unbiased — Different → Biased

Normal $N_{0,16}$

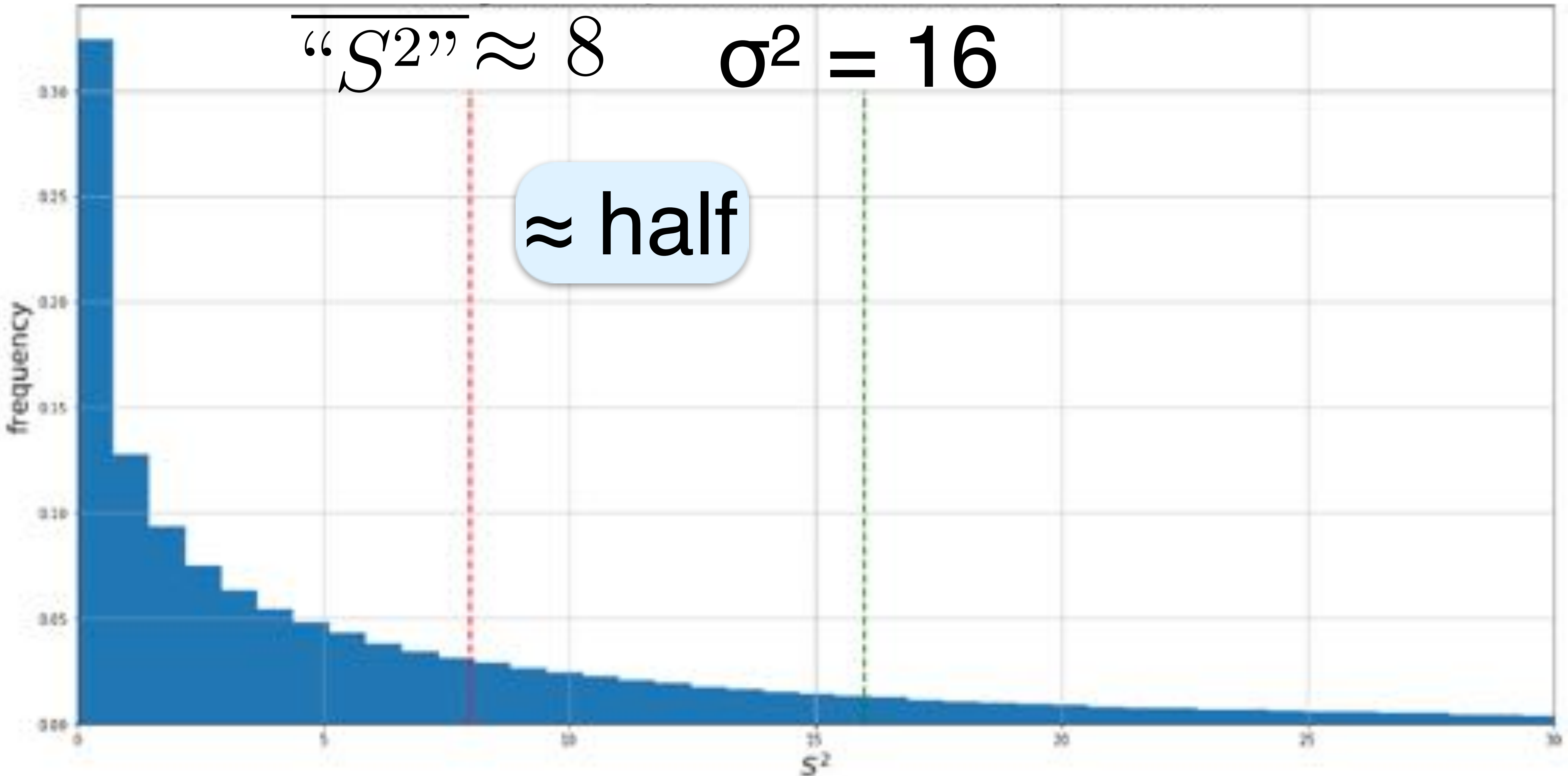
$$\text{“}S^2\text{”} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

n=2

100K Experiments

$$\overline{\text{“}S^2\text{”}} \approx E(\text{“}S^2\text{”})$$

$$E(\text{“}S^2\text{”}) \stackrel{?}{=} \sigma^2$$

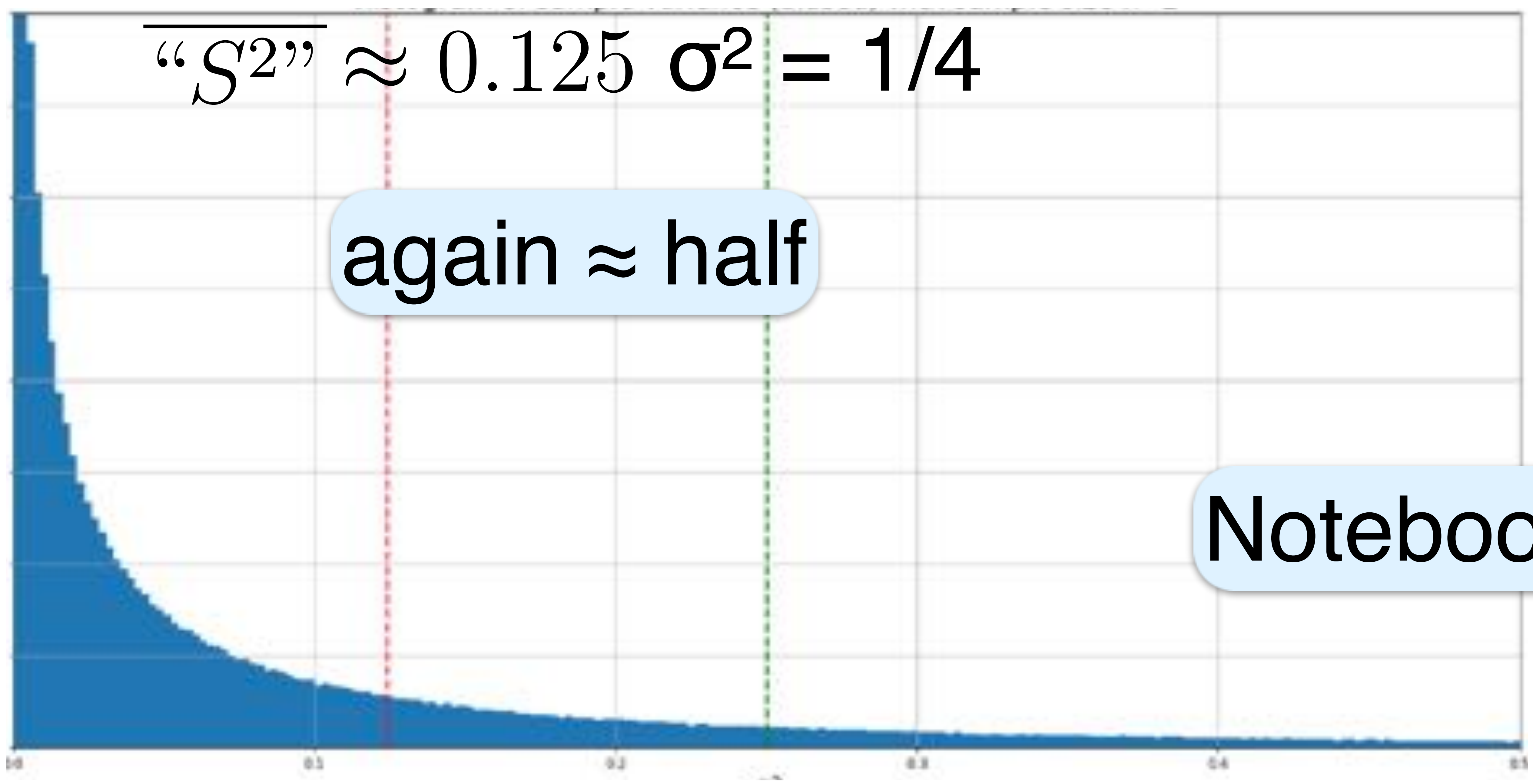


Notebook

Exponential E_2

n=2

100K Experiments



Is it bad

Program?

Numbers?

Luck?

Determine exact difference?

Bernoulli

B_p $P(1) = p$ $P(0) = 1\text{-}p = q$ $\sigma^2 = pq$

n=2 x_1, x_2 $\bar{x} = \frac{x_1+x_2}{2}$ $``S^2"(x_1, x_2) = \frac{1}{2}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2)$

x_1,x_2	$P(x_1,x_2)$	$\bar{x}$	"s^2"
0,0	q^2	0	$\frac{1}{2}((0-0)^2 + (0-0)^2) = 0$
0,1	qp	½	$\frac{1}{2}((0-\frac{1}{2})^2 + (1-\frac{1}{2})^2) = \frac{1}{2}\cdot(\frac{1}{4}+\frac{1}{4}) = \frac{1}{4}$
1,0	pq	½	¼
1,1	p^2	1	0

$$E(``S^2") = \sum_{x_1,x_2} p(x_1,x_2)\cdot ``S^2"(x_1,x_2)$$
$$= q^2\cdot 0 + qp\cdot\frac{1}{4} + pq\cdot\frac{1}{4} + p^2\cdot 0 = \frac{pq}{2} = \frac{\sigma^2}{2}$$

!

(P)review

$n = 2$ Simulation $N_{0,16}$ and E_2 $E(\text{“}S^2\text{”})$ $\neq \sigma^2$ $\approx \frac{1}{2}\sigma^2$

Exact calculation B_p

Indeed $E(\text{“}S^2\text{”}) \neq \sigma^2$

$E(\text{“}S^2\text{”}) = \frac{1}{2}\sigma^2$ Exactly Hope!

Other distributions? Other n? Why?

General n Simulations $E(\text{“}S^2\text{”}) \approx \frac{n-1}{n} \cdot \sigma^2$

Variance Estimation

(It's Almost Elementary)

Natural estimator

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Two calculations

Two-

One-pass

Bias?

Simulation

Analytically

$$E = \frac{n-1}{n}\sigma^2$$