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?

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	$\frac{1}{n}\sum_{i=1}^{n} X_i$

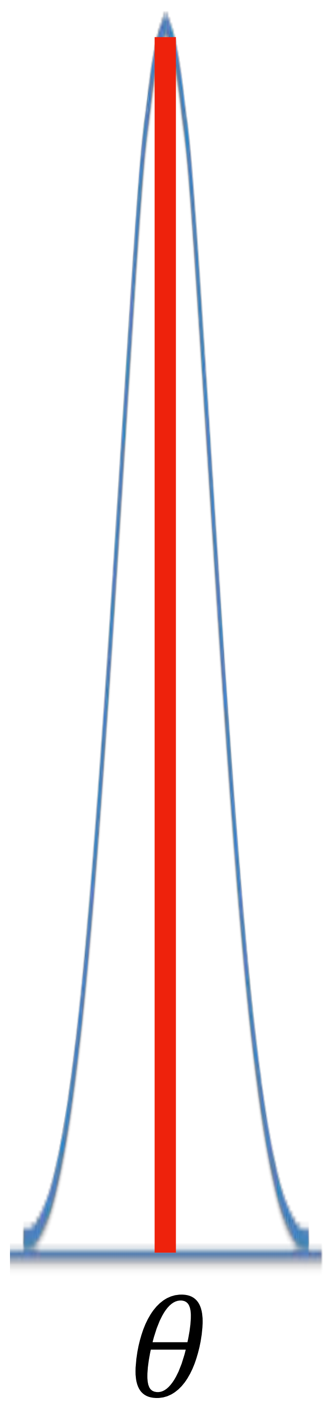
Simple — If sample is whole population, exact

Sometimes works well — Even for small samples

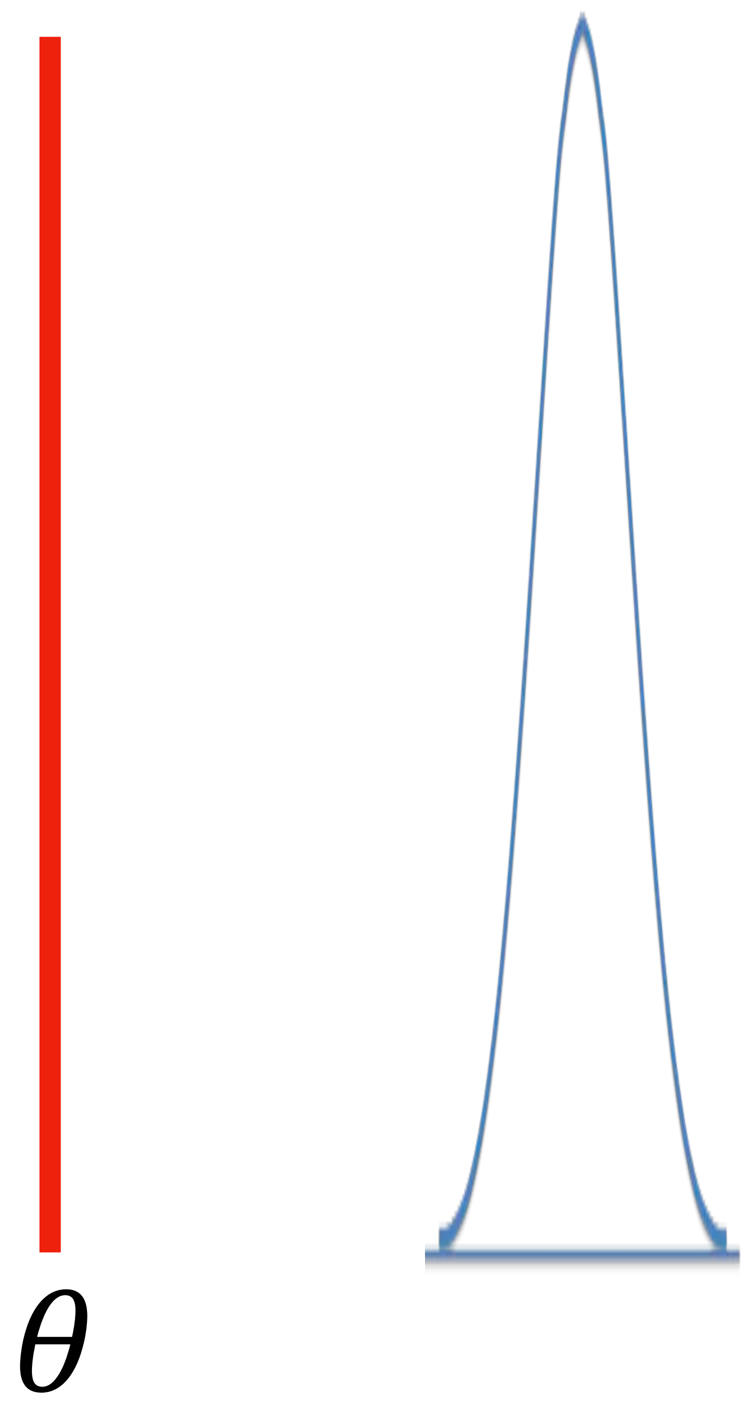
Estimator Evaluation

Parameter may have several estimators

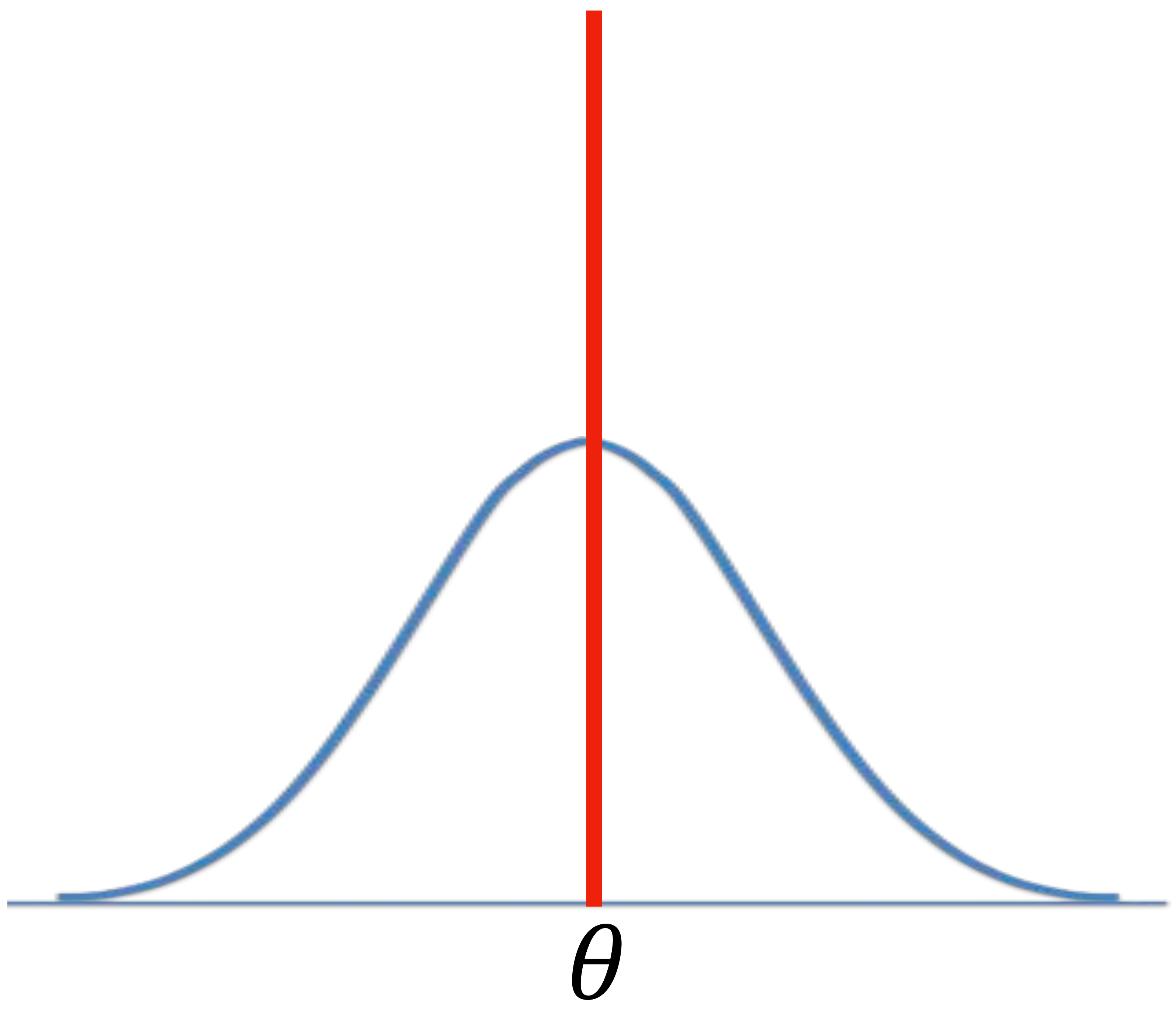
Evaluate quality of estimator for a parameter



Good



Bias



Variance

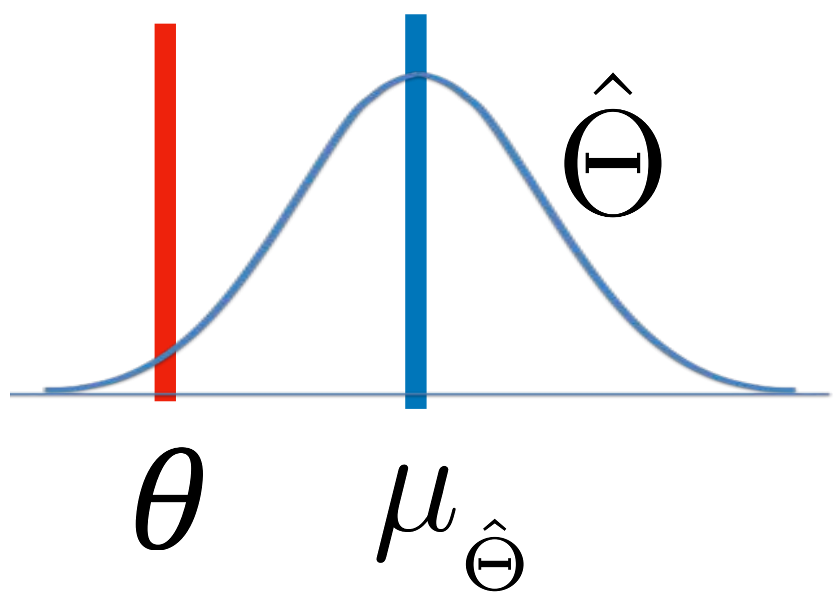
Bias

$\hat{\Theta}$ estimator for θ

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

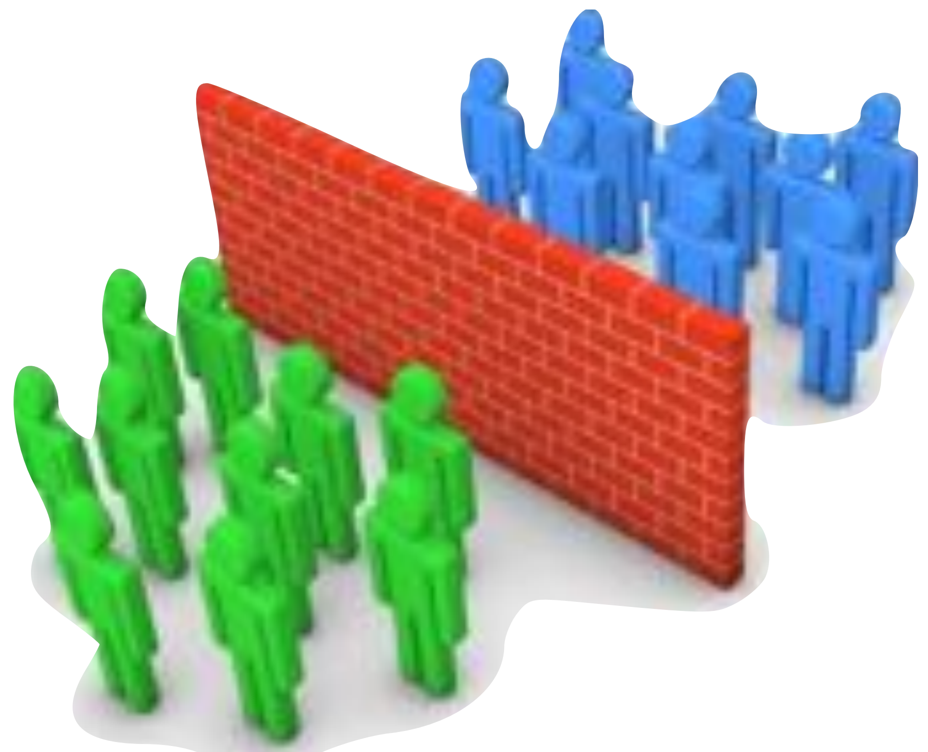
$$\mathrm{Bias}_\theta(\hat{\Theta}) \stackrel{\mathrm{def}}{=} E(\hat{\Theta} - \theta) = \mu_{\hat{\Theta}} - \theta$$

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



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

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

Sample mean is unbiased estimator for distribution mean

Sample Mean - Variance

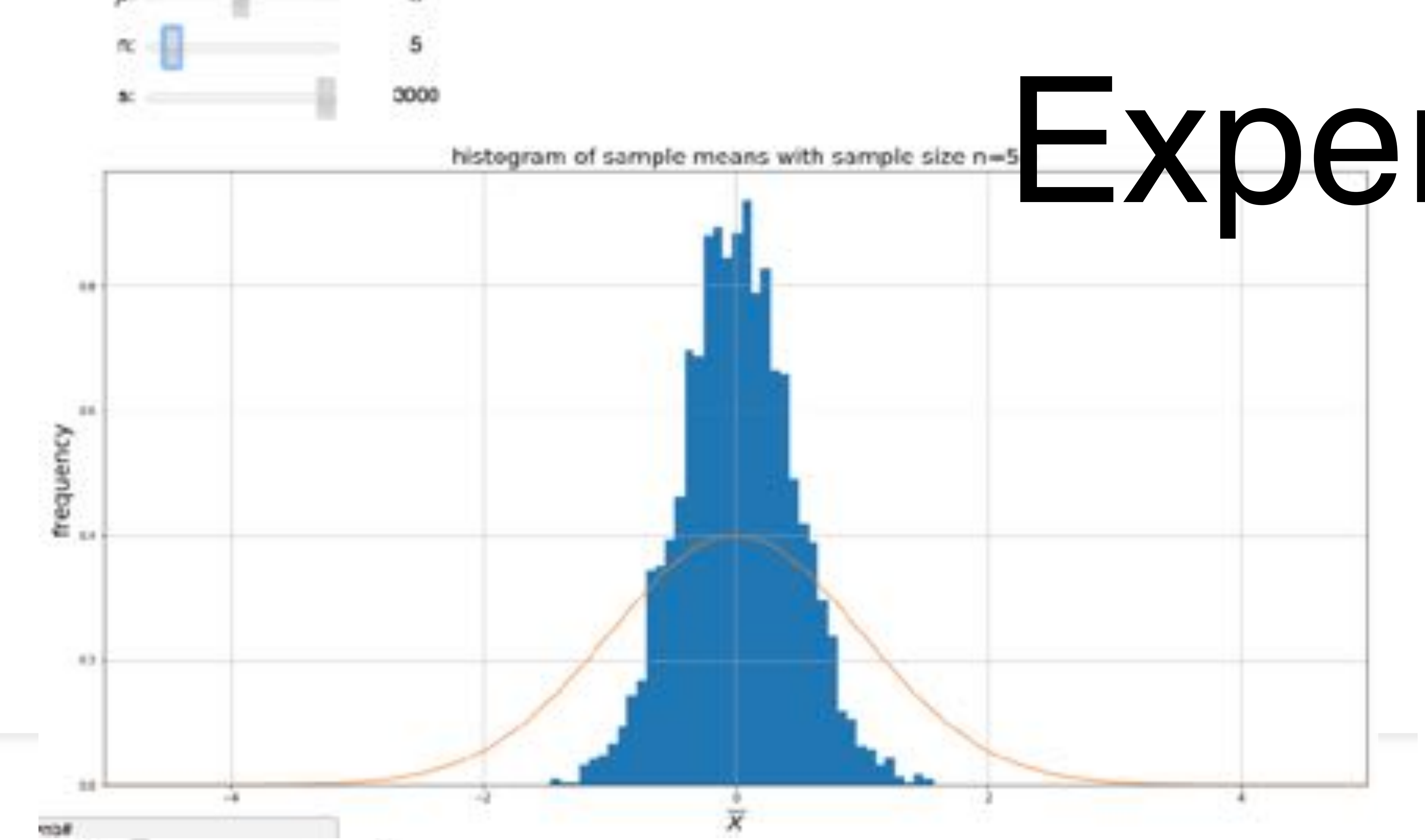
$$
\begin{aligned}
V(\overline{X}) &= V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) \\
&\overset{\perp\!\!\!\perp}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

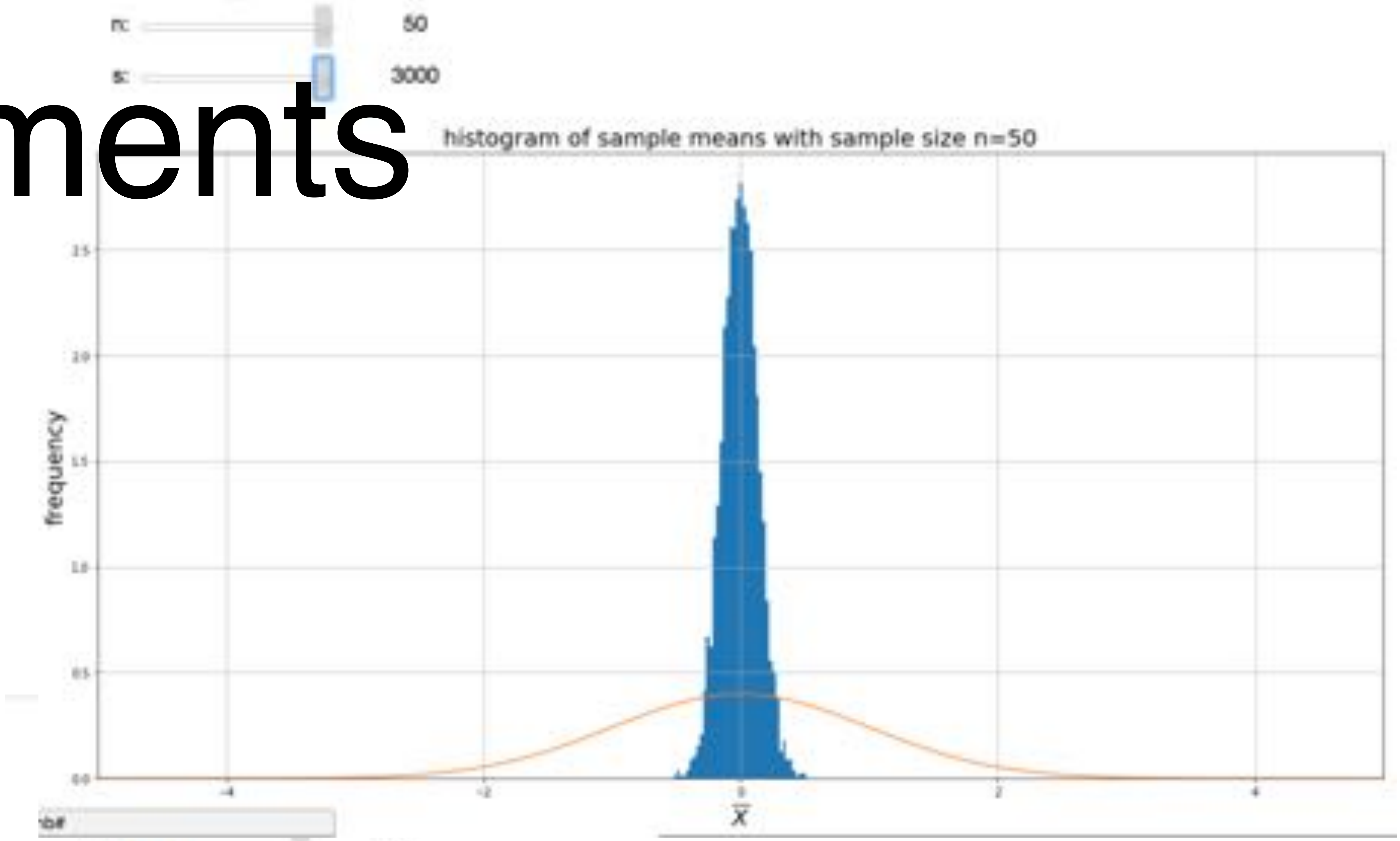
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

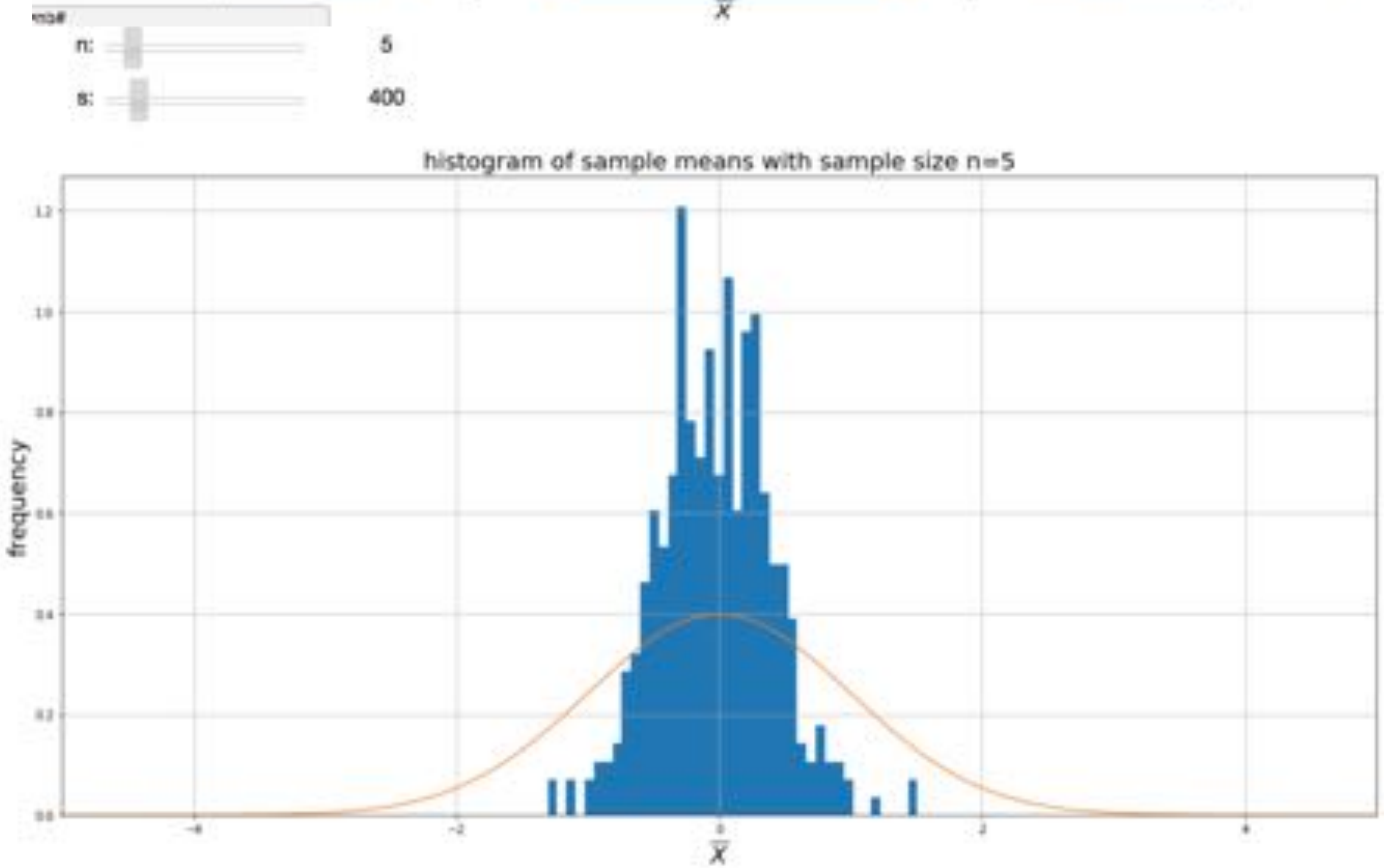
Increases with σ

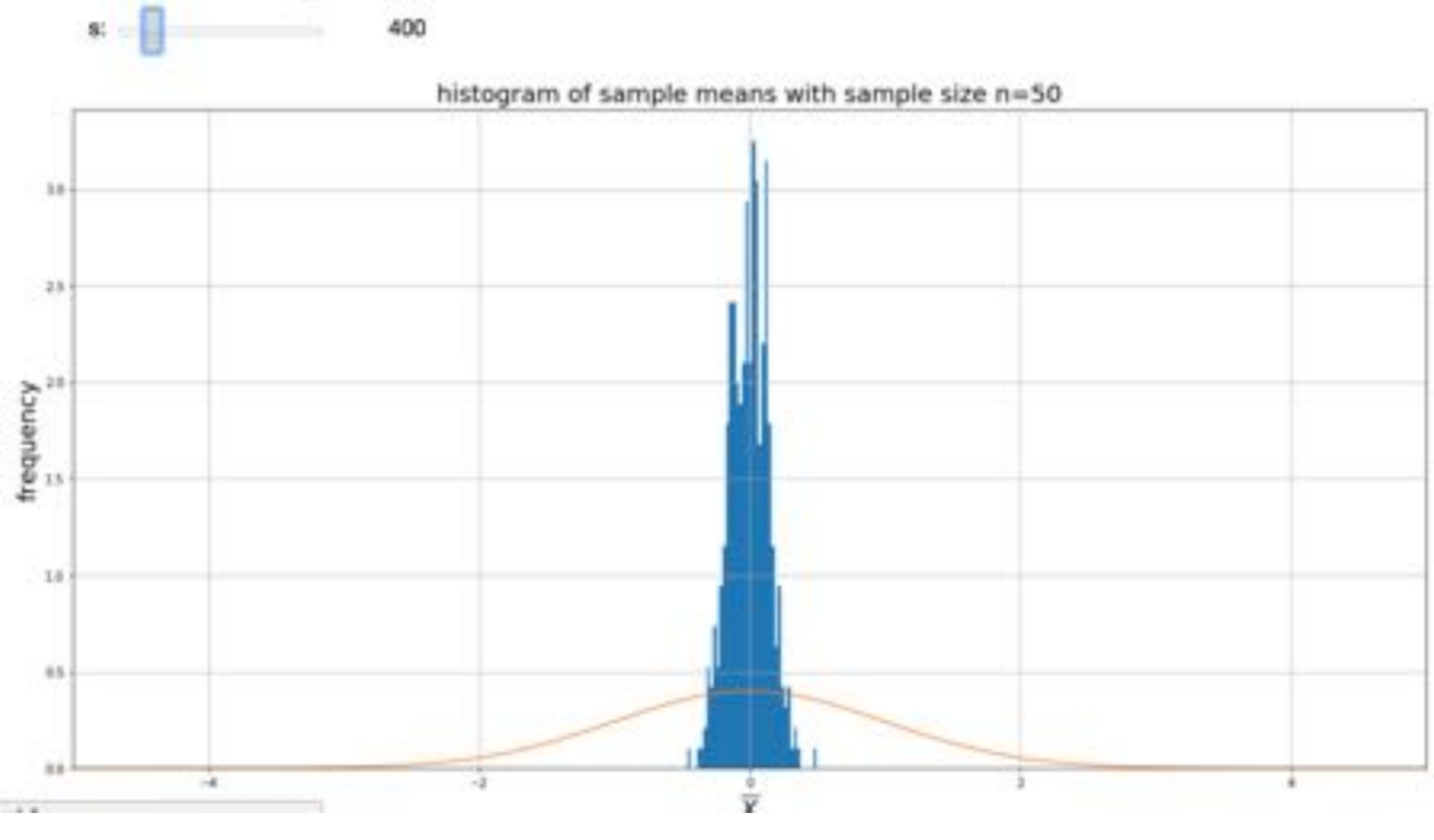
Decreases with n

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

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

Common in science and engineering

Communication

Transportation

Production

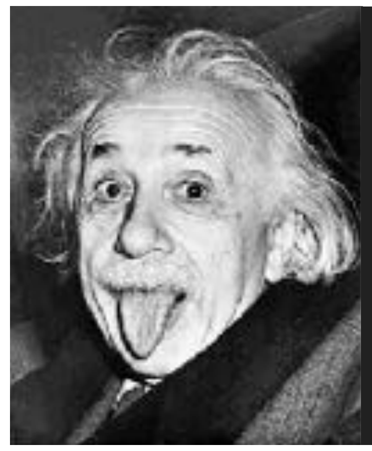
Need to re-evaluate?

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

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 $\overline{X} = 3$

$$\text{“}S^2\text{”} = 3.2$$

One pass
$$\begin{aligned}\text{“}S^2\text{”} &= \frac{1}{5}\sum_{i=1}^{5} x_i^2 - \overline{x}^2 \\ &= \frac{4+1+16+4+36}{5} - 3^2 \\ &= 12.2 - 9 \\ &= 3.2 \quad ✅\end{aligned}$$

Even more interesting…

“S²” Biased?

Mean	E → average of samples	Unbiased	Last lecture WLLN
$E(X_i) = \mu$	$\overline{X} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} X_i$	$E(\overline{X}) = \mu$	
Variance	E → average of samples	Unbiased	Stands to reason
$E(X_i - \mu)^2 = \sigma^2$	$“S^2” \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2$	? $E(“S^2”) = \sigma^2$	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 $N_{0,16}$

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

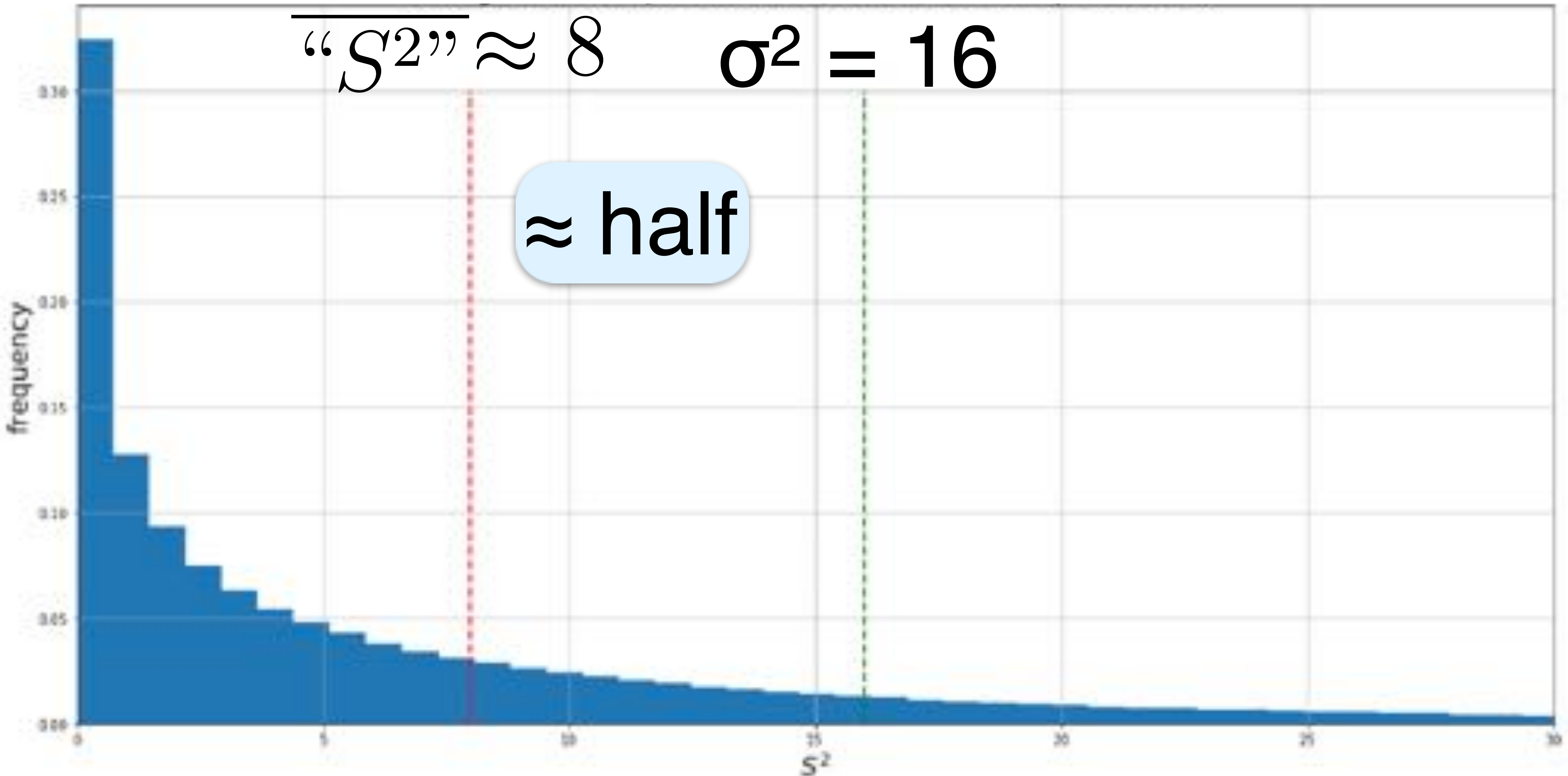
n=2

100K Experiments $\quad \overline{\text{“}S^2\text{”}} \approx E(\text{“}S^2\text{”})$

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

$\overline{\text{“}S^2\text{”}} \approx 8$ $\quad \sigma^2 = 16$

≈ half

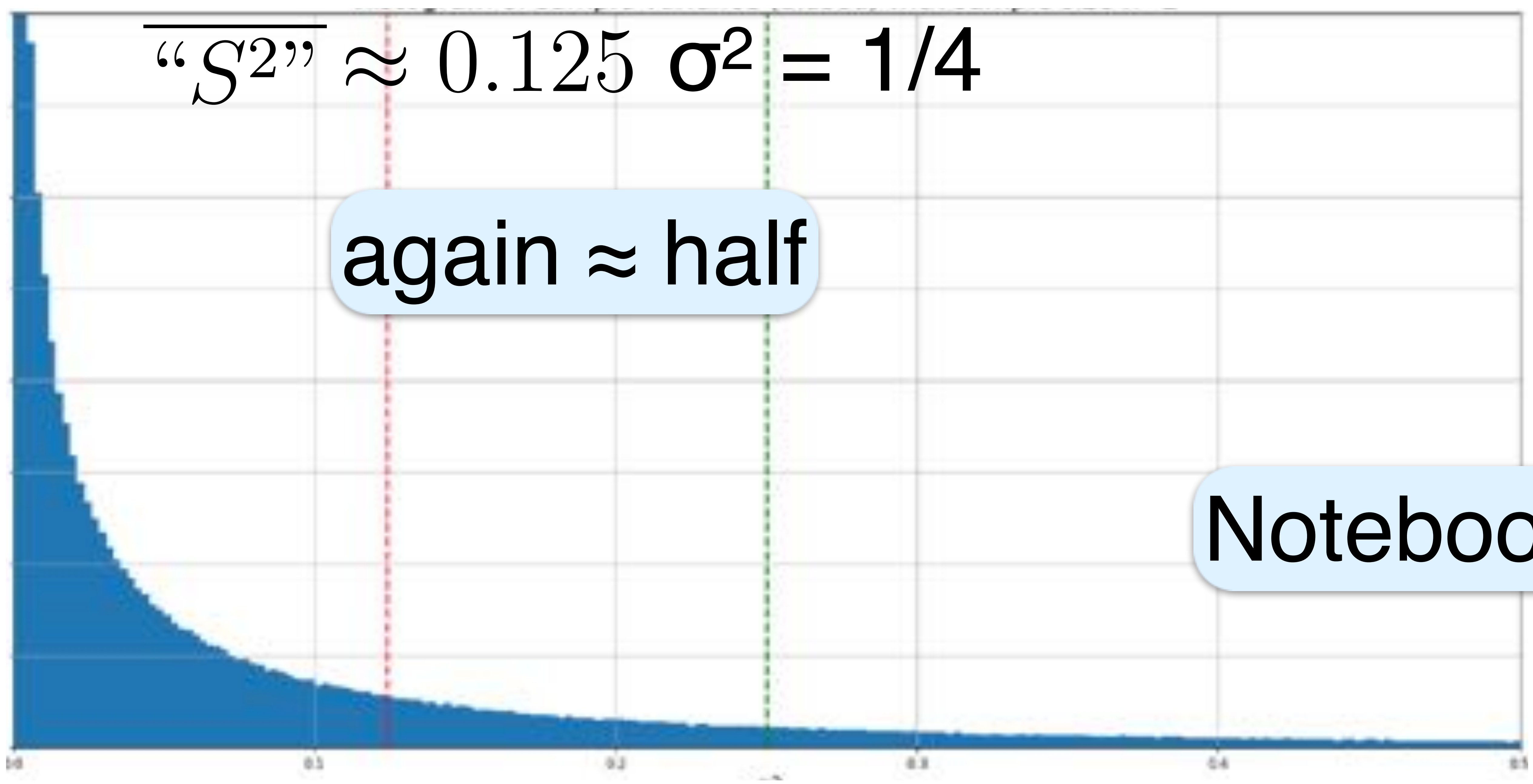


Notebook

Exponential E_2

n=2

100K Experiments



$\overline{\text{“}S^2\text{”}} \approx 0.125$ $\sigma^2 = 1/4$

again ≈ half

Notebook

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 $\overline{x} = \frac{x_1+x_2}{2}$ $“S^2”(x_1, x_2) = \frac{1}{2}((x_1 - \overline{x})^2 + (x_2 - \overline{x})^2)$

x_1,x_2	$P(x_1,x_2)$	$\overline{x}$	“s^2”
0,0	q^2	0	$\frac{1}{2}\left((0-0)^2 + (0-0)^2\right) = 0$
0,1	qp	½	$\frac{1}{2}\left((0-\frac{1}{2})^2 + (1-\frac{1}{2})^2\right) = \frac{1}{2} \cdot (\frac{1}{4} + \frac{1}{4}) = \frac{1}{4}$
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("S^2") $\neq \sigma^2$ $\approx \frac{1}{2} \sigma^2$

Exact calculation B_p

Indeed E("S^2") $\neq \sigma^2$

E("S^2") = $\frac{1}{2} \sigma^2$ 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 $E = \frac{n-1}{n}\sigma^2$

Unbiased Variance Estimation

(The mystery of the missing man)

Evaluate bias

Understand behavior

Unbiased estimator

Resolve mystery

Dispel half-truth

Mystery of the Missing Man

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

"Raw" sample variance $\text{“}S^2\text{”} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

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

Mystery

Mean

Height of 10 people

Add

Normalize by

10

9

Variance

Show $E(\text{“}S^2\text{”}) = \frac{n-1}{n} \cdot \sigma^2$

Why

How to fix

Partial Explanation

$\text{“}S^2\text{”} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ Show $E(\text{“}S^2\text{”}) = \frac{n-1}{n}\cdot\sigma^2$ “S²” under-estimates σ²

Given n points $x_1,\ldots,x_n$ $\sum_{i=1}^{n}(x_i - a)^2$ minimized for $a = \frac{x_1+\ldots+x_n}{n}$

1, -1 $(1-a)^2 + (-1-a)^2 \;= 2 + 2a^2$ minimized for a=0 average

$\sigma^2 \stackrel{\text{def}}{=} E(X-\mu)^2$ μ ≈ average of observations, not exactly

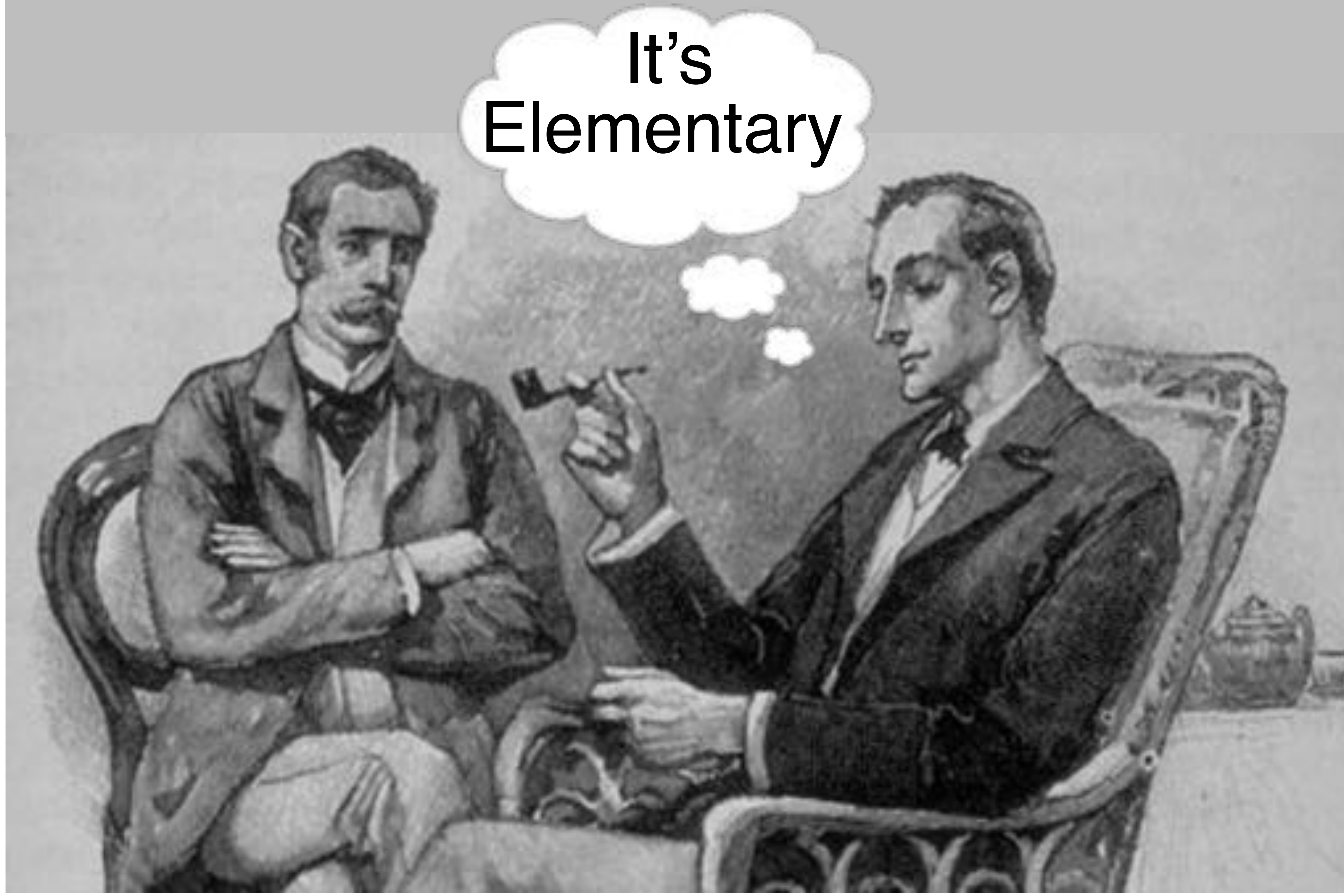
$\text{“}S^2\text{”} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ $\overline{X}$ is exact average Lower sum

Explains “S²” under-estimates σ²

Does not Explain $\frac{n-1}{n}$ Nor capture whole reason

$$
\begin{aligned}
E(\text{“}S^2\text{”}) &= E\left(\frac{1}{n}\overbrace{\sum_{i=1}^{n}}^{\text{complex}}(X_i - \overline{X})^2\right) \\
&\overset{\text{LoE}}{=} \frac{1}{n}E\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right) \\
&\overset{\text{LoE}}{=} \frac{1}{n}\sum_{i=1}^{n}E(X_i - \overline{X})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}E(X_1 - \overline{X})^2 \\
&= E(X_1 - \overline{X})^2
\end{aligned}
$$

Intuitive Simple **Elementary**



Easier understand explain

Recall: Bernoulli

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

n=2 x_1, x_2 $\overline{x} = \frac{x_1+x_2}{2}$ $\text{“}S^2\text{”}(x_1, x_2) = \frac{1}{2}((x_1 - \overline{x})^2 + (x_2 - \overline{x})^2)$

x_1,x_2	$P(x_1,x_2)$	$\overline{x}$	“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

Could get unwieldy!

$$E(\text{“}S^2\text{”}) = \sum_{x_1,x_2} p(x_1,x_2)\cdot \text{“}S^2\text{”}(x_1,x_2)$$

$$= q^2\cdot 0 + qp\cdot\frac{1}{4} + pq\cdot\frac{1}{4} + p^2\cdot 0 = \frac{pq}{2} = \frac{\sigma^2}{2}$$

!

Bernoulli 

B_P $P(1) = p$ $P(0) = 1-p = q$ $\sigma^2 = E(X-\mu)^2 = p(1-p) = pq$

Simplified calculation

$n=2$ X_1, X_2

x_1, x_2	$p(x_1, x_2)$	$\overline{x}$	$(x_1 - \overline{x})^2$
0,0	q^2	0	0
0,1	qp	½	¼
1,0	pq	½	¼
1,1	p^2	1	0

$$E(\text{“}S^2\text{”}) = E(X_1 - \overline{X})^2$$

$$= \sum_{x_1, x_2} p(x_1, x_2) \cdot (x_1 - \overline{x})^2$$

$$= 2 \cdot pq \cdot \frac{1}{4} = \frac{1}{2}pq = \frac{1}{2}\sigma^2$$ ✅

Simpler Easier to analyze

Simplified Formulation

Want to show $E(\text{“}S^2\text{”}) = \frac{n-1}{n} \cdot \sigma^2$

Asymmetric, unclear

$\stackrel{\text{def}}{=} E(X_1 - \mu)^2 \qquad X_1 \sim p$

$\stackrel{\text{def}}{=} E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = E(X_1 - \overline{X})^2$

Show $E(X_1 - \overline{X})^2 = \frac{n-1}{n} \cdot E(X_1 - \mu)^2$

Symmetric, shows difference

Simplistic Argument

$\overline{X}$ includes X_1, hence closer than μ

Doesn’t explain $\frac{n-1}{n}$

Not whole story

First n=2

General n

n=2 $E(X_1-\overline{X})^2 = \frac{1}{2}\cdot E(X_1-\mu)^2 = \frac{\sigma^2}{2}$

De-couple X_1 from $\overline{X}$ $\quad X_1-\overline{X} = X_1 - \frac{X_1+X_2}{2} = \frac{X_1-X_2}{2}$

$$E(X_1-\overline{X})^2 = E(\tfrac{X_1-X_2}{2})^2 = \tfrac{1}{4}\cdot E(X_1-X_2)^2$$

$X_2 \perp\!\!\!\perp X_1$ If difference was just from correlation between X_1 and $\overline{X}$

we would get $\frac{1}{4}\cdot E(X_1-\mu)^2 = \frac{\sigma^2}{4}$. Even smaller!

Not whole story. Randomness of X_2 reverses half of decrease. Show $E(X_1-X_2)^2 = 2\cdot E(X_1-\mu)^2$

$$E(X_1-\overline{X})^2 = \underbrace{\tfrac{1}{4}}_{\text{gain ¼ from proximity}}\cdot E(X_1-X_2)^2 = \tfrac{1}{4}\cdot \underbrace{2}_{\text{lose 2 for randomness}}\cdot E(X_1-\mu)^2 = \frac{\sigma^2}{2}$$

$$E(X_1 - X_2)^2 = 2 \cdot E(X_1 - \mu)^2$$

$$E(X_1 - X_2) = \mu - \mu = 0 \qquad E(X_1 - \mu) = \mu - \mu = 0$$

Both 0-mean

For 0-mean random variable Z

$$E(Z^2) = V(Z)$$

$$E(X_1 - X_2)^2 = 2 \cdot E(X_1 - \mu)^2 \longleftrightarrow V(X_1 - X_2) = 2 \cdot V(X_1)$$

$$V(X_1 - X_2) \overset{\perp\!\!\!\perp}{=} V(X_1) + V(X_2) = 2 \cdot V(X_1)$$

Summary for n=2

$$\left.\begin{aligned} E(\text{“}S^2\text{”}) &\overset{\text{def}}{=} E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right) \\ &= E(X_1-\overline{X})^2 \end{aligned}\right\} \text{any n}$$

$$X_1-\overline{X} = \frac{X_1-X_2}{2}$$

$$= E\left(\frac{X_1-X_2}{2}\right)^2$$

$$\overset{\text{LoE}}{=} \frac{1}{4}\cdot E(X_1-X_2)^2$$

¼ from $\overline{X}$ being closer than μ to X_1

$$\overset{\text{0-mean}}{=} \frac{1}{4}\cdot V(X_1-X_2)$$

$$\overset{\perp\!\!\!\perp}{=} \frac{1}{4}\cdot (V(X_1)+V(X_2))$$

$$\overset{\text{iid}}{=} \frac{1}{4}\cdot 2\cdot V(X_1)$$

2 from $\overline{X}$ being random

$$= \frac{1}{4}\cdot 2\cdot \sigma^2 \quad = \frac{\sigma^2}{2}$$

1/2 together

General n

$$
\begin{aligned}
E(``S^2") &= E\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\Big) \\
&= E(X_1-\overline{X})^2 \\
&= E\Big(\frac{n-1}{n}\Big(X_1-\frac{X_2+\ldots+X_n}{n-1}\Big)\Big)^2 \\
&\overset{\text{LoE}}{=} \Big(\frac{n-1}{n}\Big)^2 \cdot E\Big(X_1-\frac{X_2+\ldots+X_n}{n-1}\Big)^2 \\
&\overset{\text{0-mean}}{=} \Big(\frac{n-1}{n}\Big)^2 \cdot V\Big(X_1-\frac{X_2+\ldots+X_n}{n-1}\Big) \\
&\overset{\perp\!\!\!\perp}{=} \Big(\frac{n-1}{n}\Big)^2 \cdot \Big[V(X_1)+V\Big(\frac{X_2+\ldots+X_n}{n-1}\Big)\Big] \\
&\overset{\text{iid, var. scaling}}{=} \Big(\frac{n-1}{n}\Big)^2 \cdot \Big[\sigma^2+\frac{\sigma^2}{n-1}\Big] \\
&= \Big(\frac{n-1}{n}\Big)^2 \cdot \frac{n}{n-1} \cdot \sigma^2 \\
&= \frac{n-1}{n} \cdot \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
X_1-\overline{X} &= X_1-\frac{X_1+\ldots+X_n}{n} \\
&= \frac{(n-1)X_1-X_2-\ldots-X_n}{n} \\
&= \frac{n-1}{n}\Big(X_1-\frac{X_2+\ldots+X_n}{n-1}\Big)
\end{aligned}
$$

$\left(\frac{n-1}{n}\right)^2$ as $\overline{X}$ closer than μ to X_1

$\frac{n}{n-1}$ from $\overline{X}$ being random

$\frac{n-1}{n}$ together

Unbiased Variance Estimate

“Raw” sample variance

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

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

Bessel’s Correction

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

$E(S^2) = \sigma^2$ Unbiased estimator of variance

S^2 typically called sample variance

theoretically interesting

Large sample

Small difference

ExSample

n = 5   2, 1, 4, 2, 6

Saw $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{2+1+4+2+6}{5} = 3$

“S²” = 3.2 $\boxed{\times \frac{5}{4}}$ $\boxed{\times \frac{n}{n-1}}$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1+4+1+1+9}{4} = \frac{16}{4} = 4$$

Unbiased estimate of σ²

One-pass calculation

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

Final Simulations

Unbiased Variance Estimation

(The mystery of the missing man)

Evaluate bias

Understand behavior

Unbiased estimator

Bessel Correction

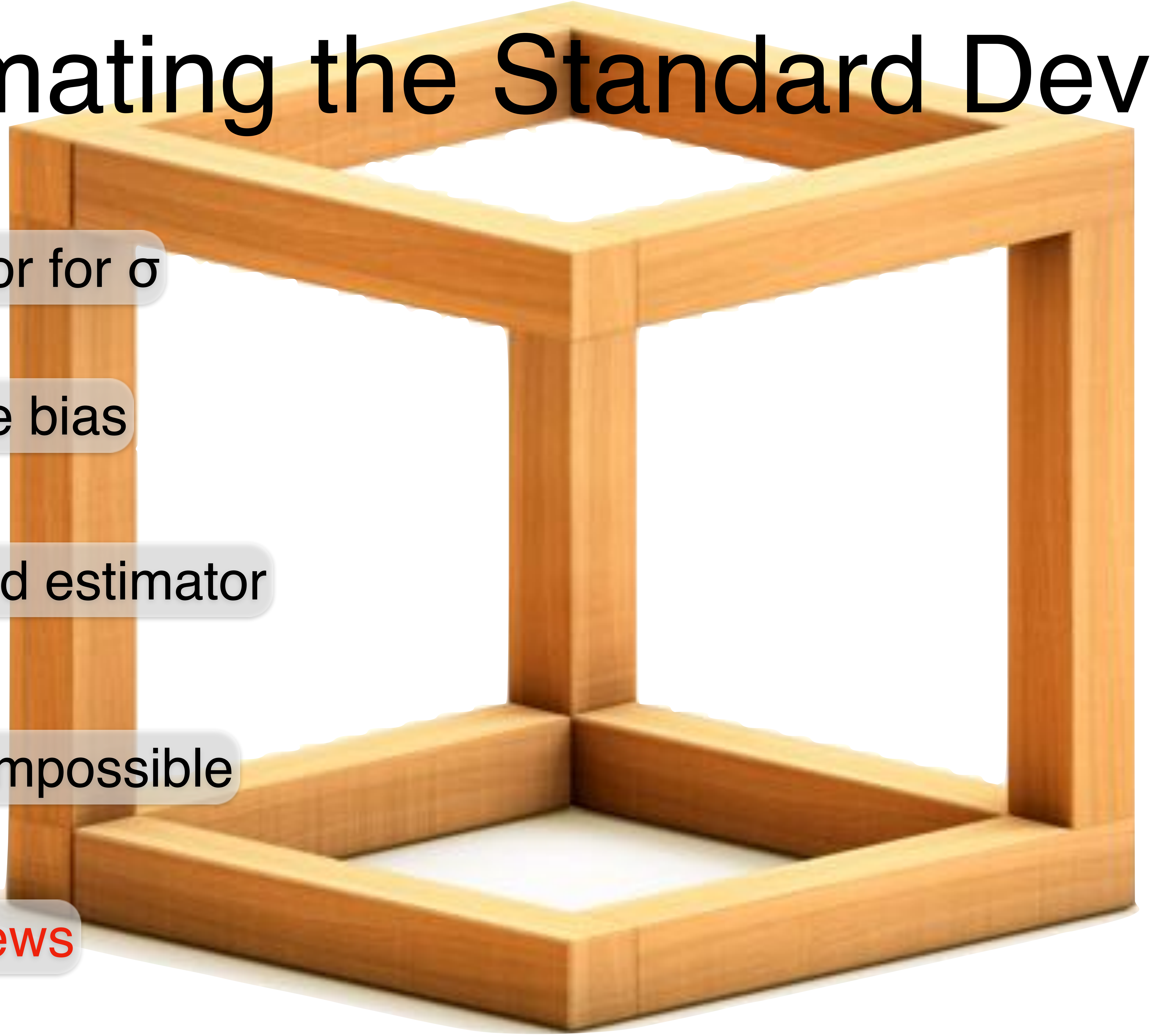
$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Resolve mystery

Dispel half-truth

Estimating σ

Estimating the Standard Deviation



Estimator for σ

Evaluate bias

Unbiased estimator

Easy x impossible

Good news

σ² → σ

Variance estimator $$S^2 \stackrel{\text{def}}{=} \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Showed $E(S^2) = \sigma^2$ S² is an unbiased estimator for σ²

Estimating σ $$\sigma = \sqrt{\sigma^2}$$

$$S \stackrel{\text{def}}{=} +\sqrt{S^2} = +\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Standard Standard-Deviation estimator

Example Evaluation Possible alternatives

ExSample

n = 5   2, 1, 4, 2, 6

Saw

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

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1+4+1+1+9}{4} = \frac{16}{4} = 4$$

Estimate for σ²

Estimate for σ

$$S = \sqrt{S^2} = \sqrt{4} = 2$$

Unbiased?

Is S an unbiased estimator for σ?

S^2 is an unbiased variance estimator

$$E(S^2) = (ES)^2 + V(S) \geq (ES)^2$$

= iff V(S)=0 iff S is a constant

$$(ES)^2 \leq E(S^2) = \sigma^2$$

$$ES \leq \sigma$$

< whenever X is not a constant

On average S underestimates σ

Concrete example

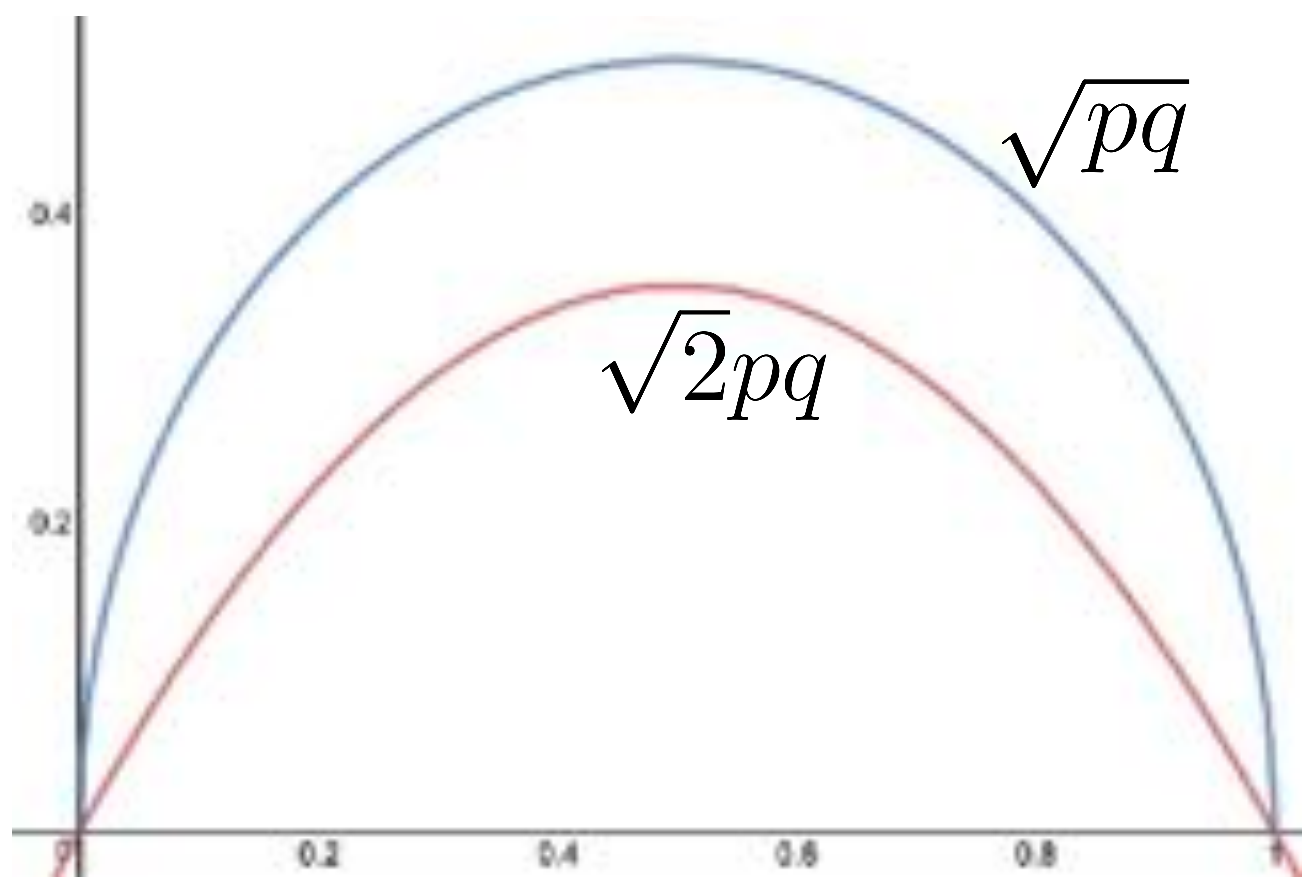
S Strictly Underestimates σ

B_p $\quad \sigma = \sqrt{p(1-p)} = \sqrt{pq}$

n=2 $\quad$ Show $\quad E(S) < \sqrt{pq}$

$$S^2 = \frac{1}{1}\left((0-\tfrac{1}{2})^2 + (1-\tfrac{1}{2})^2\right) = \frac{1}{2}$$

x_1,x_2	$P(x_1,x_2)$	$\overline{x}$	s^2	s
0,0	q^2	0	0	0
0,1	qp	½	½	$\frac{1}{\sqrt{2}}$
1,0	pq	½	½	$\frac{1}{\sqrt{2}}$
1,1	p^2	1	0	0



$$E(S) = q^2 \cdot 0 + qp \cdot \frac{1}{\sqrt{2}} + pq \cdot \frac{1}{\sqrt{2}} + p^2 \cdot 0 = \sqrt{2} \cdot pq < \sqrt{pq}$$

Unbiased Estimator for σ?

Is there an unbiased estimator for σ?

If p is known, so is σ, so nothing to estimate

Estimator must work for all distributions

For all p $\quad E(\overline{X}) = \mu \qquad E(S^2) = \sigma^2$

Is there estimator $\hat{\sigma}$ s.t. for all distributions $\quad E(\hat{\sigma}(X^n)) = \sigma$

NO $\quad$ There is no general unbiased estimator for σ !

How do you prove the impossible?

Proof Techniques

- **Proof by obviousness:** "The proof is so clear that it need not be mentioned."
- **Proof by general agreement:** "All in favor?..."
- **Proof by imagination:** "Well, we'll pretend it's true..."
- **Proof by convenience:** "It would be very nice if it were true, so..."
- **Proof by necessity:** "It had better be true, or the entire structure of mathematics would crumble to the ground."
- **Proof by plausibility:** "It sounds good, so it must be true."
- **Proof by intimidation:** "Don't be stupid; of course it's true!"
- **Proof by lack of sufficient time:** "Because of the time constrait, I'll leave the proof to you."
- **Proof by postponement:** "The proof for this is long and arduous, so it is given to you in the appendix."
- **Proof by accident:** "Hey, what have we here?!"
- **Proof by insignificance:** "Who really cares anyway?"
- **Proof by mumbo-jumbo:** $\forall \alpha \in \Phi, \exists \beta \ni \alpha * \beta = \varepsilon, \therefore \ldots$
- **Proof by profanity:** (example omitted)
- **Proof by definition:** "We define it to be true."
- **Proof by tautology:** "It's true because it's true."
- **Proof by plagarism:** "As we see on page 289,..."
- **Proof by lost reference:** "I know I saw it somewhere...."
- **Proof by calculus:** "This proof requires calculus, so we'll skip it."
- **Proof by terror:** When intimidation fails...
- **Proof by lack of interest:** "Does anyone really want to see this?"
- **Proof by illegibility:** QED
- **Proof by logic:** "If it is on the problem sheet, it must be true!"
- **Proof by majority rule:** Only to be used if general agreement is impossible.
- **Proof by clever variable choice:** "Let A be the number such that this proof works..."
- **Proof by tessellation:** "This proof is the same as the last."
- **Proof by divine word:** "...And the Lord said, 'Let it be true,' and it was true."
- **Proof by stubbornness:** "I don't care what you say- it is true."
- **Proof by simplification:** "This proof reduced to the statement 1 + 1 = 2."
- **Proof by hasty generalization:** "Well, it works for 17, so it works for all reals."
- **Proof by deception:** "Now everyone turn their backs..."
- **Proof by supplication:** "Oh please, let it be true."
- **Proof by poor analogy:** "Well, it's just like..."
- **Proof by avoidance:** Limit of proof by postponement as it approaches infinity
- **Proof by design:** If it's not true in today's math, invent a new system in which it is.
- **Proof by authority:** "Well, Don Knuth says it's true, so it must be!"
- **Proof by intuition:** "I have this gut feeling."

Handwaving	As you can see…
Induction	True for 1, 2, 3, so must be true
Example	True for this trivial example so must be true

No Unbiased σ Estimator

Even for B_p

p unknown

No unbiased estimator

No unbiased estimators for general distributions

Show for n=2 samples

Similar for any n

How do you prove the impossible?

$\hat{\sigma}$ Any estimator fo σ for B_p distributions

$\hat{\sigma}(x_1, x_2)$ Estimate of σ when observing x_1, x_2 Predetermined constants

$$
\begin{aligned}
E\left(\hat{\sigma}(X_1, X_2)\right) &= \sum_{x_1, x_2} p(x_1, x_2)\hat{\sigma}(x_1, x_2) \\
&= P(0,0)\hat{\sigma}(0,0) + P(0,1)\hat{\sigma}(0,1) + P(1,0)\hat{\sigma}(1,0) + P(1,1)\hat{\sigma}(1,1) \\
&= (1-p)^2\hat{\sigma}(0,0) + (1-p)p\hat{\sigma}(0,1) + p(1-p)\hat{\sigma}(1,0) + p^2\hat{\sigma}(1,1)
\end{aligned}
$$

Polynomial in p degree-2 polynomial

$\sigma = \sqrt{p(1-p)}$ Not a polynomial in p

The two functions differ For some p $E(\hat{\sigma}(X_1, X_2)) \neq \sigma$

Impossibility

How did we prove the impossible? Easily!

Estimators for B_p

Showed that for any estimator $\hat{\sigma}$ $E(\hat{\sigma}(X_1, X_2))$ polynomial in p

$\sigma = \sqrt{p(1-p)}$ not polynomial in p

Except: How do you prove?

For some p $E(\hat{\sigma}(X_1, X_2)) \neq \sigma$ "Don't be stupid; of course it's true!"

Therefore $\hat{\sigma}$ not unbiased

Despite joke Complete proof Give up?

Good News

Bias not so bad

Provides more freedom

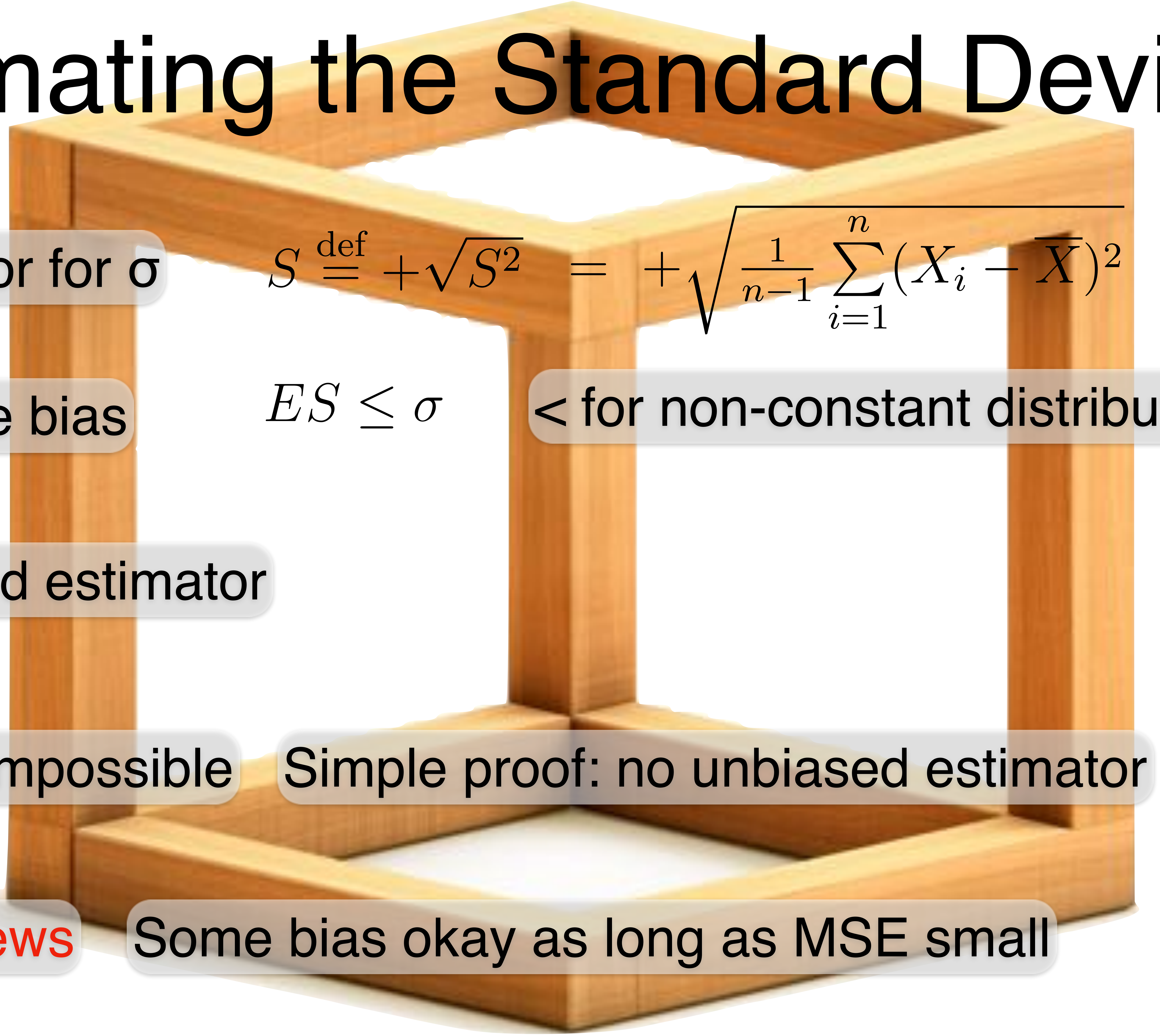
Best estimator (MSE) often biased

As the number of samples n increases

$S \rightarrow \sigma$

Consistent

Estimating the Standard Deviation



Estimator for σ

$$S \stackrel{\text{def}}{=} +\sqrt{S^2} \;=\; +\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Evaluate bias

$$ES \le \sigma$$

< for non-constant distributions

Unbiased estimator

Easy x impossible

Simple proof: no unbiased estimator

Good news

Some bias okay as long as MSE small