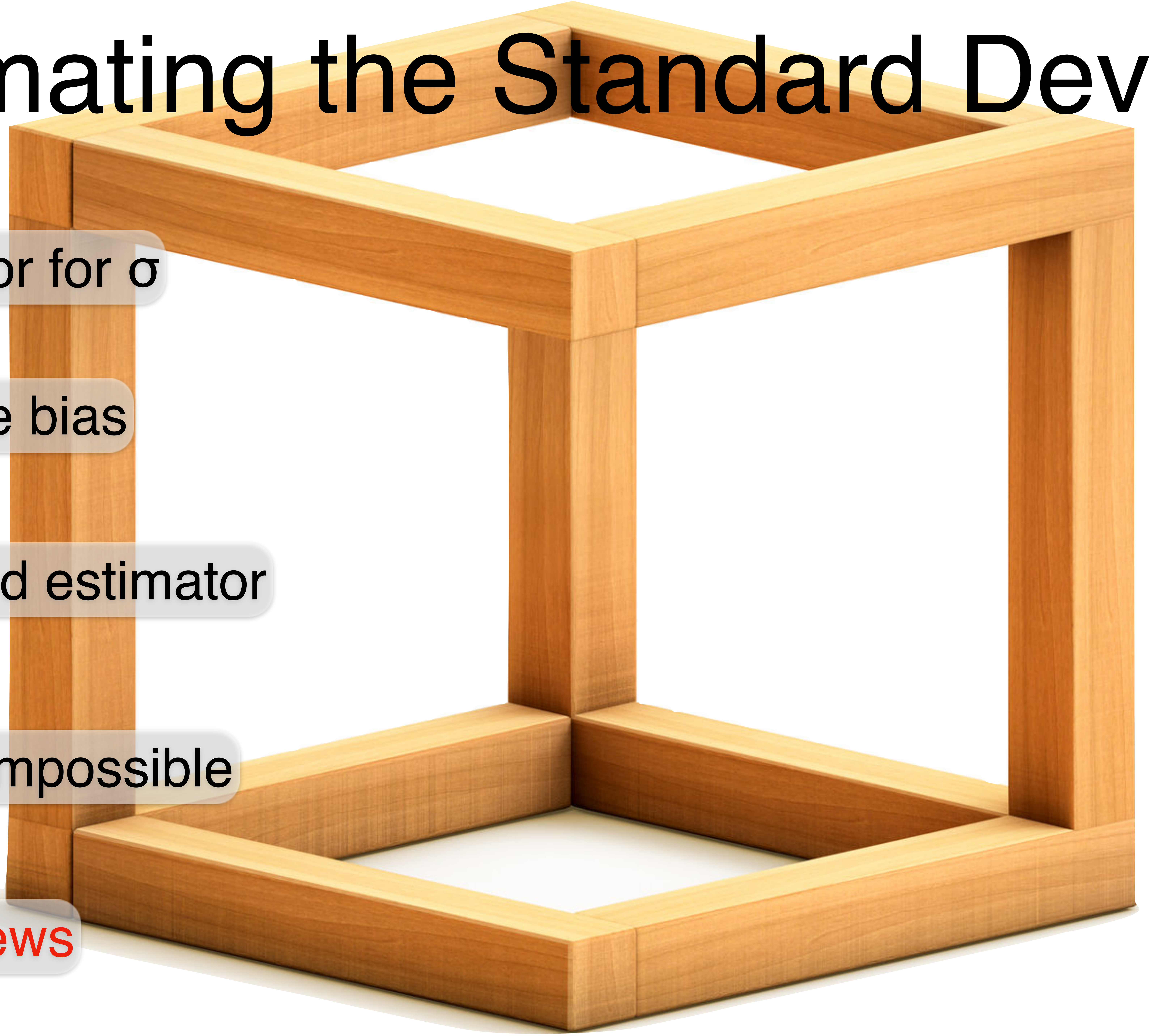
Estimating the Standard Deviation



Estimator for σ

Evaluate bias

Unbiased estimator

Easy x impossible

Good news

$\sigma^2 \rightarrow \sigma$

Variance estimator

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

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

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

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

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

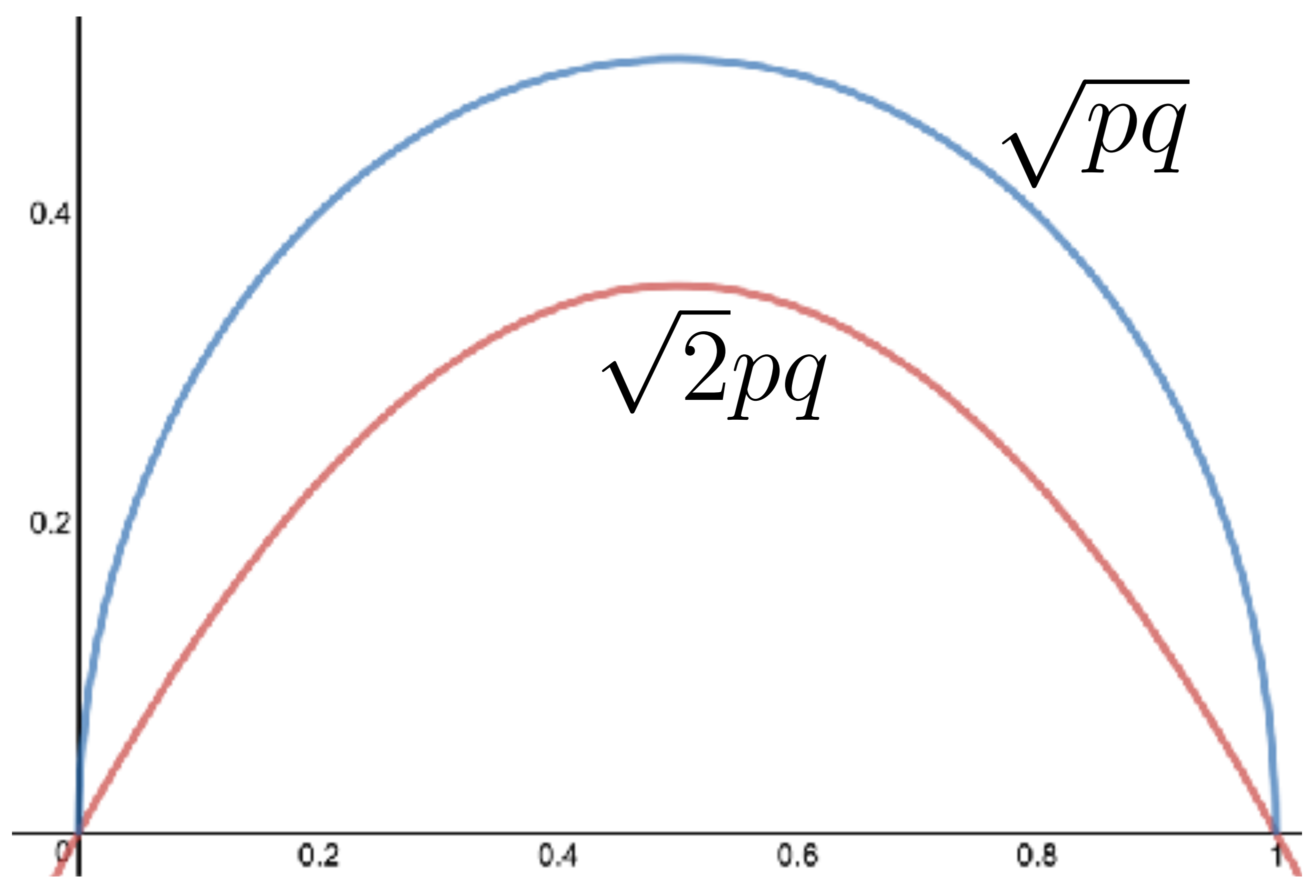
S Strictly Underestimates σ

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

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

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

x_1, x_2	$P(x_1, x_2)$	$\bar{x}$	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 $E(\overline{X}) = \mu$ $E(S^2) = \sigma^2$

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

NO 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

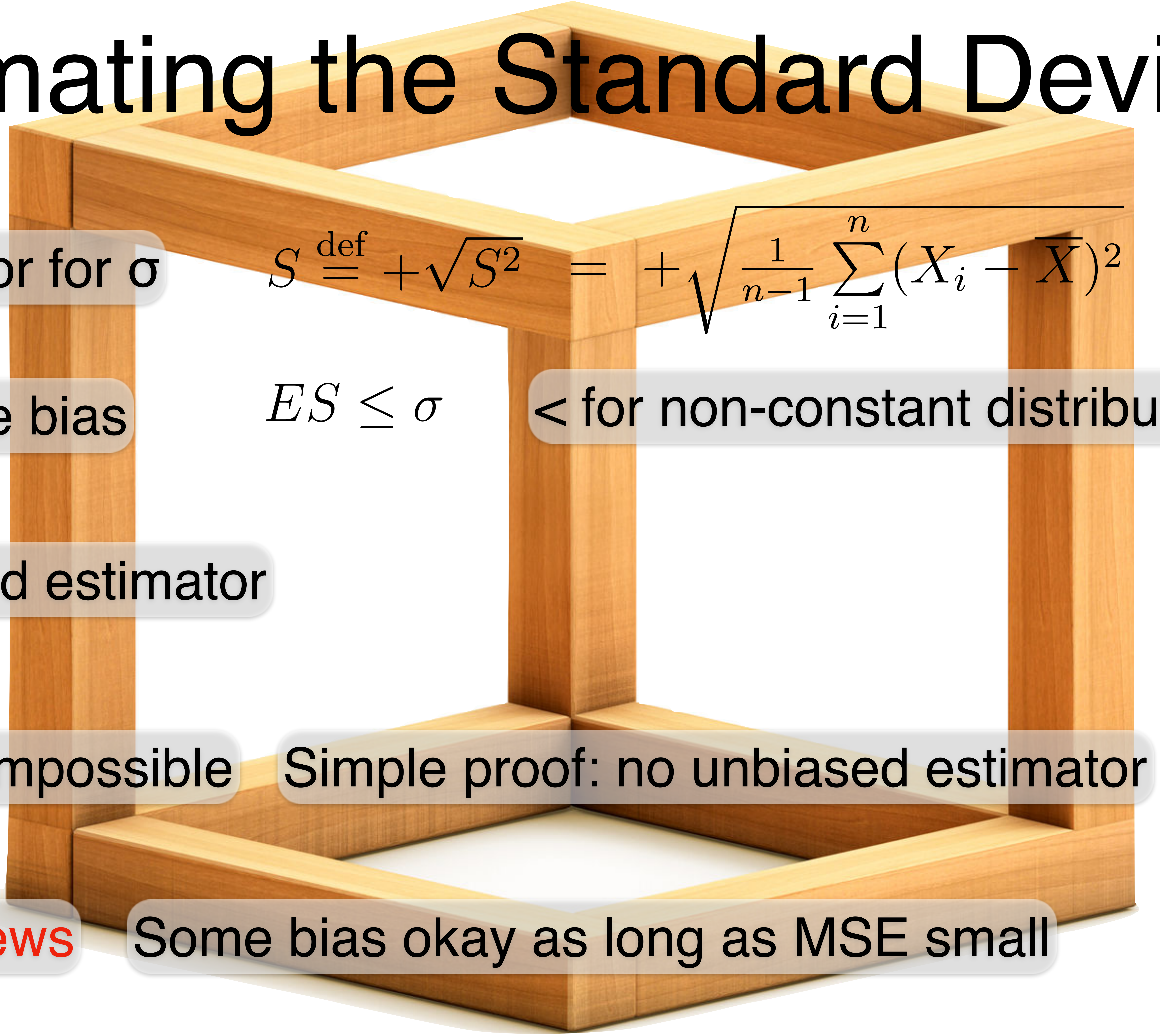
Provides more freedom

Best estimator (MSE) often biased

As the number of samples n increases $S \to \sigma$

Consistent

Estimating the Standard Deviation



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

Evaluate bias $\quad ES \le \sigma \quad$ < for non-constant distributions

Unbiased estimator

Easy x impossible $\quad$ Simple proof: no unbiased estimator

Good news $\quad$ Some bias okay as long as MSE small