Estimators and Their Properties

Gov 2001: Quantitative Social Science Methods I

Scott Cunningham

Harvard University

Spring 2026

Today's Reading

Required

- **Aronow & Miller**, §3.2.3: Estimation concepts, MSE (pp. 99–106)
- **Blackwell**, Ch. 2: Model-based inference (pp. 29–50)

Key question: What makes a good estimator?

The Three-Level Distinction

Critical vocabulary:

- **Estimand** (θ): The population quantity we want to know
  - Example: Population mean μ
  - This is a *fixed, unknown constant*
- **Estimator** ($\hat{\theta}$): A rule/formula applied to data
  - Example: Sample mean $\bar{Y} = \frac{1}{n} \sum Y_i$
  - This is a *random variable* (depends on the sample)
- **Estimate**: The number you get when you apply the estimator to your data
  - Example: $\bar{y} = 52{,}347$
  - This is a *specific number*

Estimands are targets. Estimators are procedures. Estimates are results.

Example: Voter Turnout

Research question: What fraction of eligible voters turn out?

- **Estimand**: p = true turnout rate in the population
- **Estimator**: $\hat{p} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ where $Y_i = 1$ if voter i turned out
- **Data**: Survey 1,000 voters, find 620 voted
- **Estimate**: $\hat{p} = 620/1000 = 0.62$

The estimand p is unknown. The estimate 0.62 is our best guess. The estimator tells us how to compute that guess.

What Makes a Good Estimator?

We want estimators that are:

1. **Accurate on average**: Not systematically off-target
2. **Precise**: Low variability from sample to sample
3. **Convergent**: Gets better with more data

These correspond to:

1. Unbiasedness (or low bias)
2. Low variance
3. Consistency

Bias

Definition: Bias

The **bias** of an estimator $\hat{\theta}$ for parameter θ is:

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$

Interpretation: How far off is the estimator *on average*?

- $\text{Bias}(\hat{\theta}) = 0 \Rightarrow \hat{\theta}$ is **unbiased**
- $\text{Bias}(\hat{\theta}) > 0 \Rightarrow \hat{\theta}$ tends to overestimate
- $\text{Bias}(\hat{\theta}) < 0 \Rightarrow \hat{\theta}$ tends to underestimate

Bias is about systematic error, not random error.

Example: Sample Mean is Unbiased

Claim: $\bar{Y}$ is an unbiased estimator of μ.

Proof:

$$\begin{aligned}
\mathbb{E}[\bar{Y}] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i] \\
&= \frac{1}{n} \cdot n\mu = \mu
\end{aligned}$$

Therefore: $\text{Bias}(\bar{Y}) = \mathbb{E}[\bar{Y}] - \mu = \mu - \mu = 0.$ ✓

The sample mean hits the target on average.

The Sample Variance: A Bias Story

Two candidate estimators for σ^2:

Option 1: $\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$
Option 2: $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

Which is better?

It turns out:

- $\mathbb{E}[\tilde{\sigma}^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$ (biased!)
- $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$ (unbiased)

The $n - 1$ in the denominator corrects for using $\bar{Y}$ instead of μ.

Variance of an Estimator

Definition

The **variance** of an estimator measures its spread:

$$\text{Var}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\right]$$

Interpretation: How much does $\hat{\theta}$ vary from sample to sample?

For the sample mean:

$$\text{Var}(\bar{Y}) = \frac{\sigma^2}{n}$$

Standard error: $\text{SE}(\bar{Y}) = \sqrt{\text{Var}(\bar{Y})} = \frac{\sigma}{\sqrt{n}}$

SE is the bridge to confidence intervals—it measures precision in the same units as $\bar{Y}$.

The Bias-Variance Tradeoff

Low bias, low var (Best) | Low bias, high var | High bias, low var | High bias, high var (Worst)

Target = truth. Dots = estimates from different samples.

Mean Squared Error

Definition: MSE

The **Mean Squared Error** combines bias and variance:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right]$$

The MSE Decomposition

$$\text{MSE}(\hat{\theta}) = \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$$

Proof idea: Expand $(\hat{\theta} - \theta)^2 = (\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2$ and take expectation.

MSE is the single metric that captures overall estimation error.

Why MSE Matters

Unbiased isn't everything:

Consider estimating μ with:

- $\hat{\mu}_1 = Y_1$ (just the first observation)
- $\hat{\mu}_2 = \bar{Y}$ (sample mean)

Both are unbiased! But:

- $\text{MSE}(\hat{\mu}_1) = \text{Var}(Y_1) = \sigma^2$
- $\text{MSE}(\hat{\mu}_2) = \text{Var}(\bar{Y}) = \sigma^2/n$

$\bar{Y}$ has much lower MSE for $n > 1$.

This is why we use all the data, not just one observation.

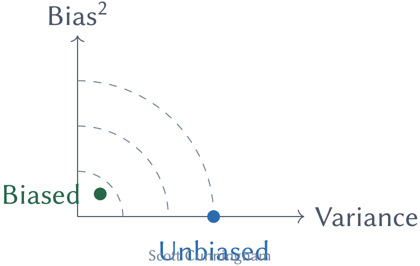
Biased But Better?

Sometimes biased estimators have lower MSE:

Political science examples:

- Small-state polls: Shrink toward national average (low *n* states are noisy)
- Election forecasts: Bayesian priors stabilize predictions at cost of some bias
- Cross-national effects: Pool countries to reduce variance of country-level estimates

The idea: Accept a little bias to get much lower variance.



Consistency (Recap)

Definition: Consistency

$\hat{\theta}_n$ is **consistent** for θ if $\hat{\theta}_n \xrightarrow{p} \theta$ as $n \to \infty$.

Intuition: With enough data, we learn the truth.

Sufficient condition: If $\text{Bias}(\hat{\theta}_n) \to 0$ and $\text{Var}(\hat{\theta}_n) \to 0$, then $\hat{\theta}_n$ is consistent.

Examples:

- $\bar{Y}$ is consistent for μ (LLN)
- $\hat{\sigma}^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$ is consistent for σ^2
- OLS coefficients are consistent under standard assumptions

Unbiased $\neq$ Consistent (A&M Theorem 3.2.16)

Common misconception: "If it's unbiased, it must converge to the truth."

Counterexample: $\hat{\mu} = Y_1$ (just the first observation)

- $\mathbb{E}[Y_1] = \mu$ ✓ (unbiased)
- $\text{Var}(Y_1) = \sigma^2$ (doesn't shrink with n!)
- Not consistent: More data doesn't help because we ignore it

The other direction: $\tilde{\sigma}^2 = \frac{1}{n}\sum(Y_i - \bar{Y})^2$

- Biased: $\mathbb{E}[\tilde{\sigma}^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$
- But consistent: $\frac{n-1}{n} \to 1$ as $n \to \infty$

Consistency requires both bias → 0 **and** variance → 0.

Hierarchy of Desirable Properties

In order of importance:

1. **Consistency**: Essential. Without it, more data doesn't help.
2. **Low MSE**: Balances accuracy and precision.
3. **Unbiasedness**: Nice to have, but not at any cost.

Summary: Unbiased $\not\Rightarrow$ Consistent, and Consistent $\not\Rightarrow$ Unbiased.

But consistent + asymptotically unbiased is typical for well-behaved estimators.

The Plug-In Principle (A&M §3.2.6)

A unifying idea: Whatever you'd compute on the population, compute on the sample.

Plug-In Estimator

Replace the population distribution with the empirical distribution of your sample.

Examples:

- Population mean $\mathbb{E}[Y] \longrightarrow$ Sample mean $\bar{Y}$
- Population variance $\text{Var}(Y) \longrightarrow$ Sample variance
- Population quantile $\longrightarrow$ Sample quantile
- $\mathbb{E}[Y|X = x] \longrightarrow$ Sample mean of Y among obs with $X = x$

LLN guarantees plug-in estimators are consistent.

Efficiency

Definition: Efficiency

Among unbiased estimators, the one with **lowest variance** is called **efficient**.

Famous result: The Cramér-Rao Lower Bound gives a minimum possible variance for unbiased estimators.

In regression: The Gauss-Markov theorem says OLS is the "Best Linear Unbiased Estimator" (BLUE).

We'll see Gauss-Markov when we get to regression.

Summary: Properties of Estimators

Property	Definition	Meaning
Unbiased	$\mathbb{E}[\hat{\theta}] = \theta$	Correct on average
Low variance	$\text{Var}(\hat{\theta})$ small	Precise
Low MSE	$\mathbb{E}[(\hat{\theta} - \theta)^2]$ small	Accurate overall
Consistent	$\hat{\theta}_n \xrightarrow{p} \theta$	Converges to truth
Efficient	Lowest variance (among unbiased)	Best in class

Remember: $\text{MSE} = \text{Bias}^2 + \text{Var}$

Key Takeaways

1. **Estimand** (target) vs. **estimator** (procedure) vs. **estimate** (number)
2. **Bias** = systematic error: $\mathbb{E}[\hat{\theta}] - \theta$
3. **Variance** = random error: $\text{Var}(\hat{\theta})$
4. **MSE = Bias**^{2} + **Variance** (the master decomposition)
5. **Consistency** is about large-sample behavior
6. **Unbiased isn't always best**—sometimes accept bias for lower variance

Next: Confidence intervals—quantifying uncertainty.

Looking Ahead

Wednesday: Confidence Intervals

- How to construct a CI using the CLT
- What a 95% CI actually means (and doesn't mean)
- Standard errors: estimated vs. known
- The t-distribution for small samples

Reading:

- A&M §3.3.1 (confidence intervals)
- Blackwell Ch. 4 (hypothesis tests—preview)