Properties of Estimators: Bias, Consistency, and Efficiency

Statistics · Module 3 — Point Estimation

Node 3.3

Position in Knowledge Graph

System:
Statistics

Structural Domain:
Point Estimation

Node:
3.3

Formal Definitions

Let $\hat{\theta}_n$ be an estimator of θ.

1. Bias

$$\mathrm{Bias}_\theta(\hat{\theta}_n) = \mathbb{E}_\theta[\hat{\theta}_n] - \theta.$$

The estimator is **unbiased** if this equals 0 for all θ.

---

2. Consistency

$$\hat{\theta}_n \xrightarrow{P} \theta \quad \text{as } n \to \infty.$$

That is, the estimator converges in probability to the true parameter.

(Strong consistency replaces convergence in probability with almost sure convergence.)

---

3. Efficiency (Introductory Notion)

Among unbiased estimators, an estimator is **more efficient** if it has smaller variance.

In asymptotic settings:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V(\theta)).$$

An estimator is asymptotically efficient if it achieves the minimal possible variance (to be formalized via Cramér–Rao bound).

Intuition

Bias measures systematic error.
Consistency measures long-run correctness.
Efficiency measures precision.

These are orthogonal dimensions:

- An estimator may be biased but consistent.
- It may be unbiased but inefficient.
- It may be consistent but converge slowly.

Inference requires balancing all three.

Structural Role

This node defines evaluation criteria for estimators.

It prepares:

- Cramér–Rao lower bound (finite-sample efficiency)
- Asymptotic normality
- Optimality theory
- Decision-theoretic comparisons

Without these properties, estimators are just formulas.

Mathematical Development

Bias–Variance Perspective

Recall:

$$\mathrm{MSE}_\theta(\hat{\theta}) = \mathrm{Var}_\theta(\hat{\theta}) + \mathrm{Bias}_\theta(\hat{\theta})^2.$$

Thus:

- Unbiasedness controls the second term.
- Efficiency controls the first term.

Consistency via Law of Large Numbers

If:

$$\hat{\theta}_n = g(\bar{X}_n),$$

and:

$$\bar{X}_n \xrightarrow{P} \mu(\theta),$$

then by continuous mapping theorem:

$$\hat{\theta}_n \xrightarrow{P} g(\mu(\theta)).$$

Thus, many estimators are consistent via LLN.

Worked Examples

Example 1: Sample Mean

If $X_i \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{E}[\bar{X}_n] = \mu,$$

$$\mathrm{Var}_\theta(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Thus:

- Unbiased
- Consistent (variance $\to 0$)
- Variance decreases at rate $1/n$

Example 2: Biased but Consistent Estimator

Let:

$$\tilde{\theta}_n = \frac{n-1}{n}\bar{X}_n.$$

Bias:

$$\mathbb{E}[\tilde{\theta}_n] - \mu = -\frac{1}{n}\mu.$$

Bias $\to 0$ as $n \to \infty$.

Thus biased but consistent.

Cross-Links

- **AI:** Bias–variance tradeoff mirrors underfitting vs overfitting.
- **Economics:** Consistency crucial for structural parameter estimation.
- **Linear Algebra:** Variance often expressible as quadratic forms.
- **Quantum:** Efficiency bounds tied to (quantum) Fisher information.

Structural Compression

Invariant:
Estimators are evaluated by expectation and variance under the model.

Mechanism:
Long-run convergence (LLN/CLT) governs consistency and asymptotics.

Cross-System Echo:
In optimization, convergence guarantees mirror statistical consistency.

Exercises

1. Prove that if $\mathrm{Var}(\hat{\theta}_n) \to 0$ and estimator is unbiased, then it is consistent.
2. Give an example of a consistent but inefficient estimator.
3. Show that MLE for Bernoulli is unbiased and consistent.
4. Explain difference between finite-sample efficiency and asymptotic efficiency.

Statistics · Module 3 — Point Estimation · Node 3.3
Concepts: bias-variance-tradeoff, convergence, maximum-likelihood-estimation
Prerequisites: 3.1, 3.2, 1.8, 1.9
Enables: 3.4, 7.1

hbar.university — own.your.cognition