Math Camp 2025 – Applied Micro

Evaluating Estimators, Convergence and Inference

Camilo Abbate & Sofia Olguin

Department of Economics, UC Santa Barbara

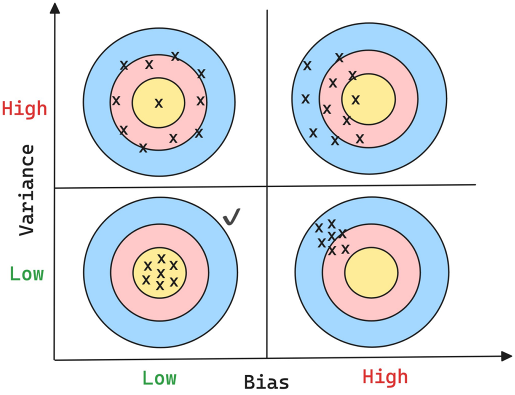
September 3, 2025

Evaluating Estimators

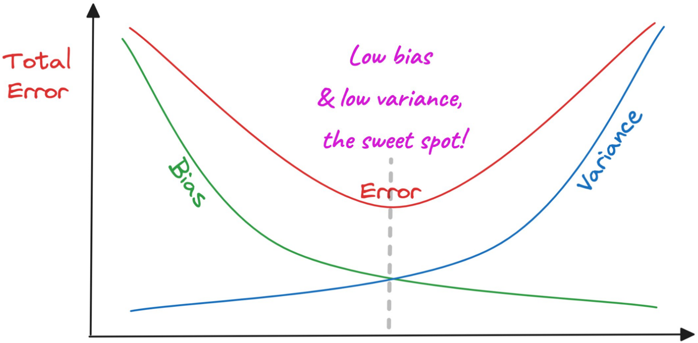
What properties should we consider when evaluating estimators?

Evaluating Estimators

What properties should we consider when evaluating estimators?



Evaluating Estimators



Bias of an estimator

Definition: Bias of an estimator $\hat{\theta}_n$ of θ:

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

Example: Variance Estimators

Two estimators for σ^2:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Example: Variance Estimators

Two estimators for σ^2:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\text{Bias}(S_n^2) = 0, \qquad \text{Bias}(\hat{\sigma}_n^2) = \frac{\sigma^2}{n}$$

$$\text{Var}(S_n^2) = \frac{2\sigma^4}{n-1}, \qquad \text{Var}(\hat{\sigma}_n^2) = \frac{2(n-1)\sigma^4}{n^2}$$

Mean Squared Error (MSE) – Key Facts

Definition:

$$\mathsf{MSE}(\hat{\theta}) = \mathbb{E}\big[(\hat{\theta} - \theta)^2\big]$$

Key Points:

- Measures the *average squared difference* between an estimator $\hat{\theta}$ and the true parameter θ.

Mean Squared Error (MSE) – Key Facts

Definition:

$$\mathsf{MSE}(\hat{\theta}) = \mathbb{E}\big[(\hat{\theta} - \theta)^2\big]$$

Key Points:

- Measures the *average squared difference* between an estimator $\hat{\theta}$ and the true parameter θ.
- Lower MSE $\Rightarrow$ estimator is *closer to the true value on average*.

Mean Squared Error (MSE) – Key Facts

Definition:

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

Key Points:

- Measures the *average squared difference* between an estimator $\hat{\theta}$ and the true parameter θ.
- Lower MSE $\Rightarrow$ estimator is *closer to the true value on average*.
- Penalizes *large errors more heavily* due to squaring.

Mean Squared Error (MSE) – Key Facts

Definition:

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

Key Points:

- Measures the *average squared difference* between an estimator $\hat{\theta}$ and the true parameter θ.
- Lower MSE $\Rightarrow$ estimator is *closer to the true value on average*.
- Penalizes *large errors more heavily* due to squaring.
- Can be decomposed into a formula involving variance and bias.

Mean Squared Error (MSE)

Show that the Mean Squared Error (MSE) can be expressed as:

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

Mean Squared Error (MSE)

Show that the Mean Squared Error (MSE) can be expressed as:

$$\mathsf{MSE}(\hat{\theta}) = \mathsf{Var}(\hat{\theta}) + \left[\mathsf{Bias}(\hat{\theta})\right]^2$$

Recall:

$$\mathsf{Var}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2].$$

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

Should we use this to decide between S_n^2 and $\hat{\sigma}_n^2$?

$$\text{Bias}(S_n^2) = 0, \qquad \text{Bias}(\hat{\sigma}_n^2) = \frac{\sigma^2}{n}$$

$$\text{Var}(S_n^2) = \frac{2\sigma^4}{n-1}, \qquad \text{Var}(\hat{\sigma}_n^2) = \frac{2(n-1)\sigma^4}{n^2}$$

Convergence in Probability

Let $\mathbf{U}_1, \mathbf{U}_2, \cdots$ be a sequence of random vectors. This sequence **converges in probability** to a random vector $\mathbf{V}$ if for any $\varepsilon > 0$:

$$\lim_{n \to \infty} P\Big(||\mathbf{U}_n - \mathbf{V}|| < \varepsilon \Big) = 1.$$

Alternatively, we write $\mathbf{U}_n \xrightarrow{p} \mathbf{V}$.

Weak Law of Large Numbers

Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ be a random sample and let $\mathbf{X}$ be a random vector with the same probability distribution as $\mathbf{X}_i$'s.

Weak Law of Large Numbers

Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ be a random sample and let $\mathbf{X}$ be a random vector with the same probability distribution as $\mathbf{X}_i$'s.
Assume that $\mathbb{E}[\mathbf{X}] < \infty$. Define $\bar{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i$. Then for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\big(||\bar{\mathbf{X}}_n - \mathbb{E}[\mathbf{X}]|| < \varepsilon\big) = 1.$$

Weak Law of Large Numbers

Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ be a random sample and let $\mathbf{X}$ be a random vector with the same probability distribution as $\mathbf{X}_i$'s.
Assume that $\mathbb{E}[\mathbf{X}] < \infty$. Define $\bar{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i$. Then for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\big(||\bar{\mathbf{X}}_n - \mathbb{E}[\mathbf{X}]|| < \varepsilon\big) = 1.$$

That is, $\bar{\mathbf{X}}_n$ converges in probability to $\mathbb{E}[\mathbf{X}]$. This is known as the **weak law of large numbers**.

More on Convergence

Suppose $Y_n \xrightarrow{p} Y$ and $Z_n \xrightarrow{p} Z$. Then

1. $cY_n \xrightarrow{p} cY$ where $c \in \mathbb{R}$
2. $Y_n + Z_n \xrightarrow{p} Y + Z$
3. $Y_n Z_n \xrightarrow{p} YZ$

More on Convergence

Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ be a random sample. Let $\hat{\theta}_n(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ be an estimator for the parameter θ, based on a sample size n. Then $\hat{\theta}_n$ is a **consistent estimator** for θ if

$$\hat{\theta}_n \xrightarrow{p} \theta$$

Central Limit Theorem

Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ be a random sample and let $\mathbf{X}$ be a random vector with the same probability distribution as $\mathbf{X}_i$'s. If $\mathbb{E}|\mathbf{X}\mathbf{X}^T| < \infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i - \mathbb{E}[\mathbf{X}]\right) \rightsquigarrow N(\mathbf{0}, \Sigma)$$

Central Limit Theorem

Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ be a random sample and let $\mathbf{X}$ be a random vector with the same probability distribution as $\mathbf{X}_i$'s. If $\mathbb{E}|\mathbf{X}\mathbf{X}^T| < \infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i - \mathbb{E}[\mathbf{X}]\right) \rightsquigarrow N(\mathbf{0}, \Sigma)$$

where $\Sigma = \mathbb{E}\Big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T\Big]$ and $\rightsquigarrow$ is short-hand for "distributed in the limit."

Central Limit Theorem

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i - \mathbb{E}[\mathbf{X}]\right) \rightsquigarrow N(\mathbf{0}, \Sigma)$$

Note that from our WLLN, $\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i - \mathbb{E}[\mathbf{X}]$ will converge in probability to zero. It converges at rate $\sqrt{n}$, however, so by multiplying by $\sqrt{n}$, we "grow" this value at the same rate it "shrinks," thus ensuring we get a distribution instead of a simply zero.