Math Camp 2025 – Applied Micro

Three Crucial Concepts in Econometrics

Camilo Abbate & Sofia Olguin

Department of Economics, UC Santa Barbara

September 2, 2025

Three Crucial Concepts

Econometrics vs. Statistics

Much of statistics is concerned with descriptive tasks that ask "what is" questions. Econometrics cares about causality and causal inference that answer "what if" questions.

Three Crucial Concepts

Given a statistical model, relating a parameter of interest to an estimand is called **identification.**

Three Crucial Concepts

Given a statistical model, relating a parameter of interest to an estimand is called **identification.**

Identification deals with the ability to uniquely determine the true values of the model parameters from the available data and model structure.

Three Crucial Concepts

Given a statistical model, relating a parameter of interest to an estimand is called **identification.**

Identification deals with the ability to uniquely determine the true values of the model parameters from the available data and model structure.

Identification is concerned with whether it is theoretically possible to recover the true parameters from the data.

Estimand, Estimate, Estimator

Definition

An **estimand** is a real number, which is a function of the probability distribution of the random variables we will get to observe.

It could be for example:

Estimand, Estimate, Estimator

Definition

An **estimand** is a real number, which is a function of the probability distribution of the random variables we will get to observe.

It could be for example:

$$Y = \mu + U \text{ where } \mathbb{E}[U] = 0 \quad \Rightarrow \quad \mu = \mathbb{E}[Y]$$

Estimand, Estimate, Estimator

Definition

An **estimand** is a real number, which is a function of the probability distribution of the random variables we will get to observe.

It could be for example:

$$Y = \mu + U \text{ where } \mathbb{E}[U] = 0 \quad \Rightarrow \quad \mu = \mathbb{E}[Y]$$

Or it could be the case that:

$$Y = \mathbf{X}^T\beta + U \text{ where } \mathbb{E}\big(\mathbf{X}U\big) = \mathbf{0}, \mathbb{E}\big(\mathbf{X}\mathbf{X}^T\big) \text{ positive definite}$$

Estimand, Estimate, Estimator

Definition

An **estimand** is a real number, which is a function of the probability distribution of the random variables we will get to observe.

It could be for example:

$$Y = \mu + U \text{ where } \mathbb{E}[U] = 0 \quad \Rightarrow \quad \mu = \mathbb{E}[Y]$$

Or it could be the case that:

$$Y = \mathbf{X}^T \boldsymbol{\beta} + U \text{ where } \mathbb{E}(\mathbf{X}U) = \mathbf{0}, \mathbb{E}(\mathbf{X}\mathbf{X}^T) \text{ positive definite}$$

$$\Rightarrow \quad \boldsymbol{\beta} = \left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right)^{-1} \mathbb{E}[\mathbf{X}Y]$$

Estimand

Examples of estimands:

- The average treatment effect (ATE) in a randomized controlled trial.
- The causal effect of a policy intervention on an outcome variable.
- The long-run equilibrium level of a macroeconomic variable.

Estimation - Estimator

Definition

Proposing an estimator for an estimand is called **estimation**.

Estimation - Estimator

Definition

Proposing an estimator for an estimand is called **estimation**. It is the process of determining the values of unknown parameters (e.g., coefficients in a regression model) using sample data.

Estimation - Estimator

Definition

An **estimator** is a function of random variables we will get to observe.

Estimation - Estimator

Definition

An **estimator** is a function of random variables we will get to observe.

Examples

Estimation - Estimator

Definition

An **estimator** is a function of random variables we will get to observe.
Examples

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\widehat{\boldsymbol{\beta}}_n = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i Y_i\right)$$

Inference - Estimate

Definition

Using an estimator to infer plausible values of an estimand is called **inference**.

Inference - Estimate

Definition

Using an estimator to infer plausible values of an estimand is called **inference**. It involves drawing conclusions about a population based on sample data.

Inference - Estimate

Definition

Using an estimator to infer plausible values of an estimand is called **inference**. It involves drawing conclusions about a population based on sample data.

Definition

An **estimate** is a realized value of the estimator given a realized sample.

Inference - Estimate

Definition

Using an estimator to infer plausible values of an estimand is called **inference**. It involves drawing conclusions about a population based on sample data.

Definition

An **estimate** is a realized value of the estimator given a realized sample.

Identification should logically come prior to inference.

Inference - Estimate

Definition

Using an estimator to infer plausible values of an estimand is called **inference**. It involves drawing conclusions about a population based on sample data.

Definition

An **estimate** is a realized value of the estimator given a realized sample.

Identification should logically come prior to inference.
If we cannot recover a parameter when we know the population distribution,

Inference - Estimate

Definition

Using an estimator to infer plausible values of an estimand is called **inference**. It involves drawing conclusions about a population based on sample data.

Definition

An **estimate** is a realized value of the estimator given a realized sample.

Identification should logically come prior to inference.
If we cannot recover a parameter when we know the population distribution, we definitely cannot recover it with a sample distribution.

Summarizing

- Estimand: the quantity to be estimated

Summarizing

- Estimand: the quantity to be estimated
- Estimate: the approximation of the estimand using a finite data sample

Summarizing

- Estimand: the quantity to be estimated
- Estimate: the approximation of the estimand using a finite data sample
- Estimator: the method or formula for arriving at the estimate for an Estimand

Let's get into **estimation**

Recall: Proposing an estimator for an estimand is called **estimation**

Let's get into **estimation**

Recall: Proposing an estimator for an estimand is called **estimation**

Let's cover two methods/processes of determining the values of unknown parameters using sample data.

Let's get into **estimation**

Recall: Proposing an estimator for an estimand is called **estimation**

Let's cover two methods/processes of determining the values of unknown parameters using sample data.

- **Maximum Likelihood**
- **Method of Moments**

Maximum Likelihood Method

Motivating problem

Let's estimate the success probability θ in a Bernoulli distribution:

- Sample: $X_1, \ldots, X_4 \sim \text{Bernoulli}(\theta)$
- Let's assumme that θ can only take two values: $1/3$ or $2/3$

Example: Observed Data

Suppose we observe: $(x_1, \ldots, x_4) = (1, 0, 1, 1)$

Let's ask the question: Which value of θ makes the data more likely?

Example: Observed Data

Suppose we observe: $(x_1, \ldots, x_4) = (1, 0, 1, 1)$

Let's ask the question: Which value of θ makes the data more likely?

If $\theta = 1/3$

$$\begin{aligned} P_1 &= (1/3)^3 \times (2/3) \\ &= 0.0247 \end{aligned}$$

If $\theta = 2/3$

$$\begin{aligned} P_2 &= (2/3)^3 \times (1/3) \\ &= 0.0988 \end{aligned}$$

Conclusion

Data are 4 times more likely with $\theta = 2/3$
$\Rightarrow$ We estimate $\hat{\theta} = 2/3$

Extension of the Example

What if θ can be $\frac{1}{3}$, $\frac{2}{3}$ or $\frac{3}{4}$?

For $\theta = 3/4$

$$\begin{aligned} P_3 &= (3/4) \times (1/4) \times (3/4) \times (3/4) \\ &= 27/256 = 0.10547 \end{aligned}$$

New conclusion

$P_3 > P_2 > P_1$

$\Rightarrow$ We estimate $\hat{\theta} = 3/4$

General Formulation - Discrete Case

For any random sample $(X_1, \ldots, X_n)$:

$$\begin{aligned} L(\theta) &= P_\theta(X_1 \ldots, X_n) \\ &= \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} \\ &= \theta^{n\bar{X}}(1-\theta)^{n(1-\bar{X})} \end{aligned}$$

General Formulation - Discrete Case

For any random sample $(X_1, \ldots, X_n)$:

$$\begin{aligned} L(\theta) &= P_\theta(X_1 \ldots, X_n) \\ &= \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} \\ &= \theta^{n\bar{X}}(1-\theta)^{n(1-\bar{X})} \end{aligned}$$

For the Bernoulli case:

$\hat{\theta}_{ML} = \bar{X}$

Formal Definition

Definition (Likelihood Function)

$$L_n(\theta) = f(X_1, \ldots, X_n; \theta)$$

Definition (Maximum Likelihood Estimator)

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} L_n(\theta)$$

- $L_n(\theta)$ is a random function of θ
- If X_i i.i.d.: $L_n(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$

Log-Likelihood

Definition (Log-Likelihood)

$$l_n(\theta) = \log L_n(\theta)$$

Key property

Since log is strictly increasing:

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} L_n(\theta) = \arg\max_{\theta \in \Theta} l_n(\theta)$$

Example: Normal Distribution

- $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$
- $\theta = (\mu, \sigma)$

Example: Normal Distribution

- $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$
- $\theta = (\mu, \sigma)$

Find the MLE for μ and σ.

Remember that the PDF of a normal distribution is:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Estimating μ

Estimator for μ

$$\hat{\mu}_{ML} = \overline{X}$$

Estimating σ

Solution

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

Final Result

Theorem (MLE for Normal Distribution)

For $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$*:*

$$\hat{\mu}_{ML} = \overline{X}$$

$$\hat{\sigma}_{ML} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Verification

We can verify that the second derivative is negative at $\hat{\sigma}_{ML}$

$\Rightarrow$ It's truly a maximum

Summary

- ML method chooses the parameter that makes data most probable
- Works for both discrete and continuous variables
- Log-likelihood is usually easier to maximize
- MLE has good theoretical properties (consistency, efficiency)

Maximum Likelihood

Theorem (Invariance Property)

If $\hat{\theta}$ is the MLE for θ, then for any function $\tau(\theta)$, the MLE for $\tau(\theta)$ is $\tau(\hat{\theta})$.

Method of Moments

Definition

Let $X_1, \cdots, X_n$ be a sample from a population with parameters $\theta_1, \cdots, \theta_k$. The **method of moments** estimator $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$ is the solution to:

$$\begin{aligned} m_1 &= M_1(\hat{\theta}_1, \cdots, \hat{\theta}_k) \\ m_2 &= M_2(\hat{\theta}_1, \cdots, \hat{\theta}_k) \\ &\vdots \\ m_k &= M_k(\hat{\theta}_1, \cdots, \hat{\theta}_k) \end{aligned}$$

where $m_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$ and $M_j = \mathbb{E}[X^j]$.

Suppose we have a random sample $X_1, \cdots, X_n$ from a normal distribution $N(\mu, \sigma^2)$. Note that $\mathbb{E}[X] = \mu$ and $\mathbb{E}[X^2] = \sigma^2 + \mu^2$. Find the method of moments estimator (MME) of μ and σ^2.

Interesting fact about MLE and MME

We just found that if the data generating process is Normal, the MLE and MME coincide.

Interesting fact about MLE and MME

We just found that if the data generating process is Normal, the MLE and MME coincide.

Does this generalize?

MLE and MME

For any **full-rank exponential family** with as many parameters as moments matched, the MLE and MM estimates coincide.
Examples:

$$\text{Poisson}(\lambda)$$
$$\text{Binomial}(p)$$
$$\text{Exponential}(\lambda)$$

But NOT in general. For many distributions outside the exponential family, or if you match fewer moments than parameters, the MM and MLE estimators will most likely be different.