Econometrics for Climate Change

Session 1 — Probability, Statistics & Linear Regression

John Gómez-Mahecha

2026-03-02

Course Overview

Who am I?

- John Gomez
- Environmental and Public economics
- Applied Econometrics
- Email: john.gomezmahecha@cmcc.it

What I will be doing in the course?

- Empirical/statistical modeling
- "Intro to econometrics for Climate Change"
- Four sessions
- Assist in the projects: Uncertainty that need from statistical methods to be answer or resolved

Sessions Overview:

1. Basics of econometrics
2. Basics of econometrics (cont.) + Estimation of Damage functions
3. Causal Inference in Statistics and Econometrics
4. Causal Inferences for Climate Change Economic Dynamics

What to do for succeed?

1. Short time + too many new concepts
2. Not time for the details -> enough time to make connections
3. How is this useful to discuss the consequences of Climate Change

Session 1: Basics of Econometrics

Objectives for this session

- Review core probability concepts
- Understand sampling and statistical inference
- Introduce the linear regression model
- Interpret OLS in climate-related applications

Question 1

Question

What is the main difference between probability and statistics?

A. Probability estimates parameters from data; statistics studies axioms.

B. Probability studies random outcomes given a model; statistics infers unknowns from data.

C. Probability is only for discrete variables; statistics for continuous variables.

D. They are equivalent terms.

Why Econometrics?

Probability, Statistics, and Econometrics

Concept: Probability

Probability is the mathematical theory that models uncertainty.

- Probability space:
  - Ω: Sample space,
  - $\mathcal{F}$: Sigma algebra,
  - P: Probability measures
- Random variables
- Probability Distributions
- Moment functions
- Conditional expectations

Why Econometrics?

Probability, Statistics, and Econometrics

Concept: Probability

Probability is the mathematical theory that models uncertainty.

Concept: Statistics

Statistics is the discipline that uses probability theory to learn about unknown population quantities from observed data.

- Estimation
- Sampling distribution
- Hypothesis testing
- Model fitting

Why Econometrics?

Probability, Statistics, and Econometrics

Concept: Probability

Probability is the mathematical theory that models uncertainty.

Concept: Statistics

Statistics is the discipline that uses probability theory to learn about unknown population quantities from observed data.

Concept: Econometrics

Econometrics is the application of statistical methods to economic data in order to

1. Estimate economic relationships
2. Test economic theories

Why Econometrics?

Probability

The mathematical language of randomness

How do random systems behave?

Statistics

Uses probability to construct estimators and perform inference

------

What can we infer form the data?

Econometrics

Uses statistical tools to answer economic question

What are the economic relationships and causal effects in data?

Why Econometrics? Random Variables

Use statistical methods

Predict

A causal relation between two variables

A future value of a variable

... for answering numerically a theoretical question

- Example in **Climate economics**:
  - Estimating impact of temperature on productivity
  - Measuring effect of carbon pricing
  - Quantifying emissions–growth relationship
- Superficially, seems a matter of which **Random Variables** we are looking to

Question 2

Question

Suppose we model next year's average temperature anomaly in a country as uncertain. Which of the following best represents a *random variable* in this context?

A. The actual temperature anomaly that will be observed next year.

B. The historical average temperature anomaly from 1980–2020.

C. A function that assigns a numerical temperature anomaly to each possible future state of the world.

D. A dataset containing temperature observations from multiple countries.

Random variables

An example:

Ω: {Very hot, Very cold}

$P(\text{Vey hot}) = p$, $P(\text{Vey cold}) = 1 - p$

RVs have a Value and Index structure.

A random variable will be the function:

$$T : \Omega \to \mathbb{R}$$

that assigns numerical variables to each state

$$T(\text{Very hot}) = 35$$
$$T(\text{Very cold}) = 5$$

Question 3

Question

What are the differences between the following random variables?

1. Number of floods across countries
2. CO_2 per capita across countries
3. Number of storms per year
4. Temperature anomaly over time
5. Emissions

RV - Index structure

How the variables are indexed? - Type of data

1. Time series
2. Cross sectional
3. Longitudinal or panel data
4. Pool

Time →

Unit	1	2	...	t	...	T
1	X_1	X_2		X_t		X_T

RV - Index structure

How the variables are indexed? - Type of data

1. Time series
2. Cross sectional
3. Longitudinal or panel data
4. Pool

Time

Unit

	1
1	X_1
2	X_2
...	
i	X_i
...	
n	X_n

RV - Index structure

How the variables are indexed? - Type of data

1. Time series
2. Cross sectional
3. Longitudinal or panel data
4. Pool

Time →

Unit ↓	1	2	...	t	...	T
1	X_{11}	X_{12}		X_{t1}		X_{T2}
2	X_{12}	X_{22}		X_{t2}		X_{T2}
...	...	...		...		
i	X_{1i}	X_{2i}	...	X_{ti}	...	X_{Ti}
...	...	...		...		...
n	X_{1n}	X_{2n}	...	X_{tn}	...	X_{Tn}

RV - Index structure

How the variables are indexed? - Type of data

1. Time series
2. Cross sectional
3. Longitudinal or panel data
4. Pool

Time

Unit

	1	...	T1	T1+1	...	T2
1	X_{11}	...	$X_{T_1 1}$			
...	X_{12}	...	$X_{T_1 2}$			
n_1	$X_{1 n_1}$	...	$X_{T_1 n_1}$			
n_{1+1}				$X_{(T_1+1)(n_1+1)}$	...	$X_{(T_2)(n_1+1)}$
...				...		...
n_2				$X_{(T_1+1)n_2}$	...	$X_{T_2 n_2}$

RV - Index structure

How the variables are indexed? - Type of data

1. Time series
2. Cross sectional
3. Longitudinal or panel data
4. Pool

Always identified your **unit of observation**!

RV - Value structure

Two type of values structure:

Concept: Discrete RVs

The set of values it can take (that is the **support**) is finite or countably infinite.

Examples?

Concept: Continuous RVs

It takes values in an uncountable set, typically an interval of $\mathbb{R}$.

Examples?

RV - Probability functions

Concept: Probability Function

A probability function assigns/indicates probabilities to events in a probability space $\mathcal{F}$

$$f(X = x) : \ \mathcal{F} \rightarrow [0, 1] \tag{1}$$

1. Non-negativity - Non negative probabilities
2. Normalization - Sum to one
3. Countable additivity - The probability of disjoint events is the sum of the events' probabilities

Discrete RVs

Continuous RVs

RV - Probability functions

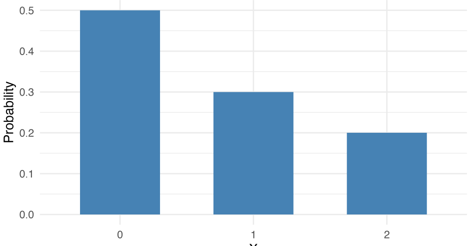
Concept: Probability Function

A probability function assigns/indicates probabilities to events in a probability space $\mathcal{F}$

$$f(X = x) : \ \mathcal{F} \to [0, 1] \tag{1}$$

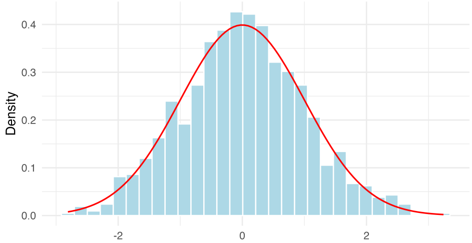
Discrete RVs

Probability <u>mass</u> function



Continuous RVs

Probability <u>density</u> function

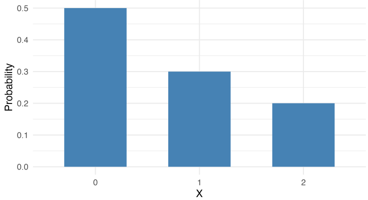


RV - Expected value

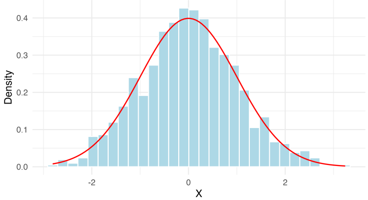
Concept: Expected value

The expected value (or **mean**) of X, denoted $E[X]$, is the probability-weighted average of its possible values.

Discrete RVs



Continuous RVs

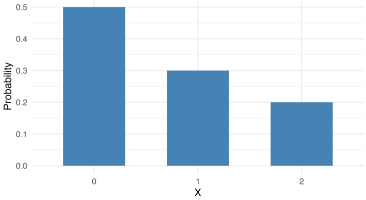


RV - Expected value

Concept: Expected value

The expected value (or **mean**) of X, denoted $E[X]$, is the probability-weighted average of its possible values.

Discrete RVs



$$E[X] = \sum_x x\, f(x)$$

Continuous RVs

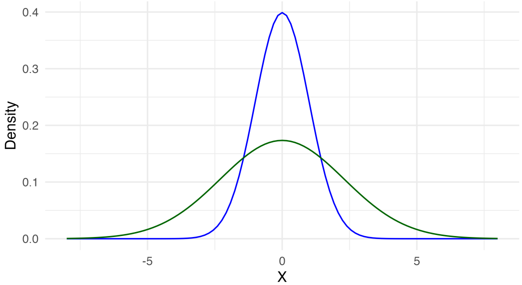
$$E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

RV - Variance

Concept: Variance

The variance of a random variable X, denoted $\mathrm{Var}(X)$, measures the expected squared deviation of X from its mean.

$$\mathrm{Var}(X) = E\Big[(X - E[X])^2\Big] \tag{2}$$



Statistic and Inference

Where does Statistics enter?

- Let us suppose we have a RV X of interest (e.g., GDP, Emissions)
- We would like to gather information about the dsitribution of this variable, lets say the mean ($E[X]$ or μ_X)
- μ_X is the value for the whole the population
- We do not have access to that!!

The best I can do is to get a **Random Sample** (S)

Statistic and Inference - Sample

Concept: Random Sample

A **random sample** of RV X is a series of n realizations of X $(x_1, x_2, ...)$ that satisfy a series of conditions:

1. Each x_i is the reult of randomly select a value for X, **independently** of the other possible values (x_j)
2. Each realization will follow the distribution of X

That is, we will work with **Independently and indetically distributed**, or **iid**, samples.

Statistic and Inference - Estimator

Based on the sample, we will build estimators

Concept: Estimator

Any function of the Sample ($Estimator(x) : S \to \mathbb{R}$, build with the purpose to learn something about the distribution of a given RV.

- We want those estimators to have a series of properties:
  - **Unbias**: $E[Estimator] = Parameter$
  - **Efficiency**: $EstA$ is more efficient than $EstB$ if it has lower variance
  - **Consistency**: $\bar{X} \to^p Parameter$
  - A known distribution
- Other properties are convenient, but not fully required:
  - **Linear**: $\bar{X}$ is a linear function of the elements in S

Question 5

Let us assume we want to estimate the expected value of the random variable X, for which we have an iid sample S of size n. We three possible candidates and we want to evaluate their properties:

	Is it lineal?	**Unbiased?**	**Efficient?**
$\tilde{X} = x_2 + x_3$			
$\hat{X} = x_1$			
$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$			

Question 4

Question

What does it mean for an estimator to be **BLUE**?

A. Best Unbiased Linear Estimator
B. Biased Linear Uniform Estimator
C. Bayesian Linear Unrestricted Estimator
D. Best Linear Unbiased Estimator

Cool property of the sample average

- We already see that the $\bar{X}$ is unbias and linear
- It can be shown, but it is also the most efficient within all linear estimators
- The **Law of large Numbers** will guarantee it is **consistent**
- The **The Central Limit Theorem** tell us that

$$\frac{\bar{X} - \mu_x}{\sigma^2/n} \sim N(0,1)$$

How we will use this for Inference?

Linear regression - Relation between two variables

Law of Iterated Expectations

$$E[Y] = E[E[Y|X]]$$

We observe a random sample:

$$\{(Y_i, X_i)\}_{i=1}^{n}$$

Climate example: Expected agricultural output given temperature

From Conditional Expectation to Regression

Define:

$$m(X) = E[Y|X]$$

Linear approximation:

$$m(X) \approx \beta_0 + \beta_1 X$$

Regression = Best Linear Predictor

Linear Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Where:

- Y_i : outcome (e.g., emissions per capita)
- X_i : regressor (e.g., GDP per capita)
- u_i : unobserved factors

OLS estimator derivation