Chapter 5: Statistics

Introduction to Probability Theory & Econometrics

Julian Klix | E600 Mathematics | Fall Semester 2026

UNIVERSITÄT MANNHEIM

Outline

In this chapter, we discuss

- A few basics from probability theory
- Random variables
- Convergence, Weak Law of Large Numbers, Central Limit Theorem

What is Econometrics?

- Econometrics = economic data analysis
  - mostly: understanding relationships between economic variables
    - Does inflation cause unemployment?
    - Does automation cause the labour share to decrease?
    - Has policy X reduced tax evasion in country Y?
- Purposes
  - Understanding economic relationships
    - ...to develop new theories
    - ...to test existing theories
    - ...to guide economic policy
  - Evaluating economic policies

Probability Spaces

Probability Space

A **probability space** $\mathcal{P}$ is a triple $\mathcal{P} := (\Omega, \mathcal{A}, P)$, together with

- Sample space Ω, the (atomic) set of possible outcomes $\omega \in \Omega$,
- Event space $\mathcal{A} \subseteq 2^{\Omega}$, the set of all possible events $A \in \mathcal{A}$ such that
  - it contains the sample space, $\Omega \in \mathcal{A}$
  - it is closed under compliments: if $A \in \mathcal{A}$, then also $A^c := \Omega \setminus A \in \mathcal{A}$
  - it is closed under (finite) unions: if $A_i \in \mathcal{A} \;\; \forall i \in \mathcal{I}$, then also $\bigcup_{\mathcal{I}} A_i \in \mathcal{A}$
- Probability measure $P : \mathcal{A} \to [0, 1]$ assigning probability $P(A)$ to $A \in \mathcal{A}$ with
  - $P(\Omega) = 1$ [and consequently $P(\emptyset) = 0$]
  - if $A, \; B \in \mathcal{A}$ and A and B are disjoint ($A \cap B = \emptyset$), then $P(A \cup B) = P(A) + P(B)$

Probability Spaces

- The event space $\mathcal{A}$ is sometimes also called the σ-Algebra of $\mathcal{P}$
- A simple probability space for some finite sample space $\Omega = \{\omega_1, \dots, \omega_n\}$
  - Sample space as given
  - Complete event space: $\mathcal{A} = 2^\Omega$ [every possible subset]
  - Laplace measure: $P(A) = \frac{|A|}{|\Omega|}$ [assuming every event equally likely]
- Example: Rolling a fair dice
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $\mathcal{A} = 2^\Omega$ e.g. $A_{<5} = \{1, 2, 3, 4\} = \{\omega \in \Omega : \omega < 5\} \in \mathcal{A}$ or $A_{\text{odd}} = \{1, 3, 5\} \in \mathcal{A}$
  - $P(A) = \frac{|A|}{6}$, e.g. $P(\{1, 2, 3, 4\}) = \frac{4}{6} = \frac{2}{3}$ and $P(\{1, 3, 5\}) = \frac{1}{2}$

$\Rightarrow$ for any given Ω, neither $\mathcal{A}$ nor $\mathcal{P}(\cdot)$ are unique!

Independence of Events

Conditional Probability

Consider two events $A, B \in \mathcal{A}$. Then, the probability of event A conditional on event B, denoted as $P(A|B)$ is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Independence of Events

Two events $A, B \in \mathcal{A}$ are stochastically independent if $P(A \cap B) = P(A) \cdot P(B)$

$\Rightarrow$ independence implies that $P(A|B) = P(A)$

$\Rightarrow$ without independence, $P(A \cap B) = P(A) + P(B) - P(A \cup B)$

Bayes' Rule

Bayes' Rule

Let $\{B_i\}_{i\in\{1,\dots,n\}} \in \mathcal{A}$ be disjoint sets where $\bigcup_i B_i = \Omega$ and $P(B_i) > 0 \; \forall i = 1, \dots, n$ (called a partition of Ω) and take another event $A \in \mathcal{A}$. Then

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_i P(A|B_i)P(B_i)}$$

- Allows to reverse conditioning from $P(A|B)$ to $P(B|A)$
- Often referred to (in Economics) as Bayesian Updating
  - Interested in B, while event A occurs (correlated with B)
  - Prior probability $P(B)$ without knowledge of A
  - Posterior probability $P(B|A)$ after "learning" about A

Random Variables

- Events are sets, prefer working with scalars, vectors,... [$\rightarrow$ **Random Variable**]
- Example for a random variable: rolling a fair dice - did we roll a 6?

$$X(\omega) = \mathcal{I}(\omega = 6) \quad \left[X(\omega) = 1 \text{ if } \omega = 6, \text{ otherwise } X(\omega) = 0 \right]$$

- Random variables allow to focus on *lower-dimensional*, relevant outcomes, e.g.
  - Ω = space of meteorological conditions, $X(\omega)$ = temperature at $\omega \in \Omega$
  - Ω = space of economic conditions, $X(\omega)$ = GDP at $\omega \in \Omega$

Random Variable and Random Vector

Consider a probability space $(\Omega, \mathcal{A}, P)$. Then, a function $X : \Omega \mapsto \mathbb{R}$ is called a random variable with realization $x \in \mathbb{R}$, and a vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ where X_i, $i \in \{1, \ldots, n\}$ are random variables, is called a random vector.

- Formally, X needs to be measurable $\rightarrow$ disregarded in this course!

Density and Distribution Functions

- Density and distribution function
  - Distribution function F_X: $F_X(x) = P(X \leq x)$
  - Density of continuous random variables: $f_X = \frac{dF_X}{dx}$, continuous s.t. $F_X(\infty) = 1$
  - Density of discrete random variables (= frequency): $f_X(x) = P(X = x)$

- Examples
  - Discrete random variable: eyes on a fair dice
    - Distribution function $F_X(t) = P(X \leq t) = \sum_{i=1}^{t} \frac{1}{6}$
    - Frequency function $f_X(x) = \frac{1}{6} \forall x \in \{1, \ldots, 6\}$
  - Continuous random variable: univariate normal distribution
    - Distribution function $\int_{\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)}{2\sigma^2}\right) dx$
    - Density function $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)}{2\sigma^2}\right)$

Expectation and Variance

- Expected value $\mathbb{E}[X]$
  - Generally: $\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx$
  - Discrete: $\mathbb{E}[X] = \sum_{i \in I} x_i \cdot f_X(x_i)$ (I: index set of possible outcomes)
  - Interpretation: probability weighted average of X

Rules for Expectations

Let $X, Y : \Omega \mapsto \mathbb{R}$ be random variables, $a, b \in \mathbb{R}$ and $g : \mathbb{R} \mapsto \mathbb{R}$. Then,

- Linearity: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- Functional Expectation $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx$
- Only for independent X, Y (defined later): $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
- Jensen's Inequality: $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ for convex functions g

Variance, Covariance and Correlation

- Variance $Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
  - Useful result: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
  - Standard deviation $sd[X] = \sqrt{Var[X]}$ [often used preferably since scaling equally]

- Covariance $Cov(X, Y)$ of two random variables X and Y
  - Definition: $Cov(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$
  - Note: $Cov(X, X) = Var(X)$

- Correlation $\rho_{X,Y}$ between two random variables X and Y
  - Definition: $\rho_{X,Y} = \frac{Cov(X,Y)}{sd(X)sd(Y)}$
  - Normalized $\rho_{X,Y} \in [-1, 1]$
    $\Rightarrow$ sign: direction of mutual movement between X and Y
    $\Rightarrow$ magnitude: strength of the relationship between X and Y

Joint Distributions

- Jointly continuous RVs $X,\ Y$
  - Joint density function $f_{X,Y} : \mathbb{R} \times \mathbb{R} \mapsto [0, \infty]$ s.t. $\int_X \int_Y f_{X,Y}(x, y) dy\, dx = 1$
  - Joint distribution function $F_{X,Y} : \mathbb{R} \times \mathbb{R} \mapsto [0, 1]$ such that

$$F_{X,Y}(x, y) = \int_0^x \int_0^y \phi_{X,Y}(u, v) dudv$$

- Jointly discrete RVs $X,\ Y$
  - Joint frequency function $p_{X,Y}$ defined as $p_{X,Y}(x, y) = P(X = x, Y = y)$
  - Joint distribution function $P_{X,Y}$ defined as $P_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ or such that

$$P_{X,Y}(x, y) = \sum_{\{x_i < x, y_j < y\}} p_{X,Y}(x_i, y_j)$$

$\Rightarrow$ Generalize univariate RV to multiple RVs with corellation structures

Marginal Distributions

- Disregard one/multiple jointly distributed RVs [here: focus on X, disregard Y]

  $\Rightarrow$ Recombine all random vectors (x, y_i) with different y_i into single x
- Jointly continuous RV $X,\ Y$
  - Combining by integration $\rightarrow$ marginal density $f_X(x)$ such that

$$f_X(x) = \int_Y f_{X,Y}(x, y)\, dy$$

- Jointly discrete RV $X,\ Y$
  - Combining by summation $\rightarrow$ marginal frequency $p_X(x)$ such that

$$p_X(x) = \sum\nolimits_Y p_{X,Y}(x, y)$$

$\Rightarrow$ proceed as with any other univariate or joint density/frequency

Independence of Random Variables

Independence of Random Variables

Two random variables $X,\ Y$ are called independent if $\forall A, B \subset \mathbb{R}$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

$\Leftrightarrow$ for the joint distribution function and $a, b \in A, B$

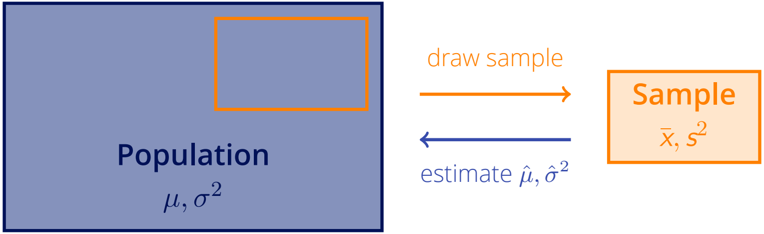
$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$$

$\Leftrightarrow$ for the joint density/frequency function

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{or } p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

$\Rightarrow$ Independence will often make your life easier, but is not always given.

Population/Distribution and Samples



- True distributions of interest but seldomly directly observable
- Sample properties to inform your beliefs of population variables ($\rightarrow$ estimation)
- Be careful what notation is refering to (e.g. $\mathbb{E}$, *Var*)! Convention:
  - Population variables: Greek letters $(\mu, \sigma, \rho, \ldots)$
  - Sample variables: Roman letters $(\bar{x}, s, r, \ldots)$
  - Estimators (for population variables): Hatted greek letters $(\hat{\mu}, \hat{\sigma}, \hat{\rho}, \ldots)$

Conditional Expectations

- Conditional expectation is either a specific value $\mathbb{E}[X|Y=y]$ or a function $\mathbb{E}[X|Y]$

Conditional Expectation (Function)

Let X, Y be two continuous random variables in the same probability space. Then

1. Given conditional density $f_{X|Y}(x|y)$, define $\mathbb{E}[X|Y=y]$, in short $\mathbb{E}[X|y]$, as

$$\mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx \quad \text{where } f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

2. There is a (measurable) function $g : \mathbb{R} \mapsto \mathbb{R} : g(Y) = \mathbb{E}[X|Y]$, that is

$$g(y) = \mathbb{E}[X|y] \quad \forall y \in Y$$

- Implications: The conditional expectation of X given Y ...

  ...is a function of Y! ...and thus, itself a random variable.

Rules for Conditional Expectations

- $\mathbb{E}[X|Y]$ for discrete RVs is defined analogously (integral $\leftrightarrow$ sum)

Rules for Conditional Expectation Functions

Let X, X_1, X_2, Y be random variables in the same probability space. Then,

- $\mathbb{E}[\alpha X_1 + \beta X_2 \,|\, Y] = \alpha \mathbb{E}[X_1 \,|\, Y] + \beta \mathbb{E}[X_2 \,|\, Y]$ for $\alpha, \beta \in \mathbb{R}$ (Linearity)
- $\mathbb{E}[\mathbb{E}[X|Y]\,] = \mathbb{E}[X]$ (Law of Iterated Expecations)
- $\mathbb{E}[\mathbb{E}[X|Y]\,|\,Y] = \mathbb{E}[X|Y]$ (Tower Property)
- $\mathbb{E}[f(Y)X|Y] = f(Y)\mathbb{E}[X|Y]$ for all (measurable) functions f
- $\mathbb{E}[X|Y] = \mathbb{E}[X]$ if X and Y are independent
- $\mathbb{E}[f(X)|Y] \geq f(\mathbb{E}[X|Y])$ for all convex functions f (Jensen's Inequality)

Convergence in Probability Spaces

Convergence in Distribution

A series of RVs $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution to some RV X, written $X_n \xrightarrow{d} X$ iff

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad \forall x \text{ at which } F_X \text{ is continuous}$$

Convergence in Probability

A series of RVs $\{X_n\}_{n\in\mathbb{N}}$ converges in probability to some RV X, written $X_n \xrightarrow{P} X$ or $plim_{n\to\infty}(X_n) = X$ if

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \qquad \forall \epsilon > 0$$

Convergence in Probability Spaces

Almost Sure Convergence

A series of RVs $\{X_n\}_{n\in\mathbb{N}}$ converges almost surely to some RV X, written $X_n \xrightarrow{a.s.} X$ if

$$P\big(\lim_{n\to\infty} X_n = X\big) = 1$$

- All concepts are relatively similar $\rightarrow$ distinction here not too relevant
- Concepts presented in increasing order of strength, that is

$$X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X$$

- Useful result: Continuous mapping theorem

$$X_n \xrightarrow{a.s./P/d} X \implies g(X_n) \xrightarrow{a.s./P/d} g(X) \quad \text{for all continuous functions } g$$

Weak Law of Large Numbers

Weak Law of Large Numbers

Consider a set of independent and identically distributed (iid) RVs $\{T_i\}_{i=1,\dots,n}$ with $\mathbb{E}({T_i}^2) < \infty$. Then

$$\bar{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i \xrightarrow{P} \mathbb{E}(T_i)$$

- Interpretation: sample averages converge to population expected values.

  $\Rightarrow$ only requires fairly general assumptions

  $\Rightarrow$ can be used to build simple estimators

- A strong version exists with almost sure convergence (Kolmogorov's Theorem)

Central Limit Theorem

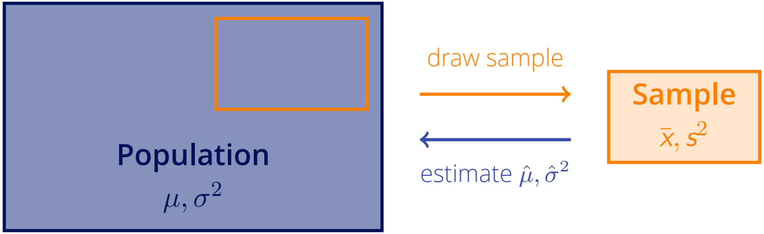
Central Limit Theorem

Consider a set of independent and identically distributed (iid) RVs $\{T_i\}_{i=1,\ldots,n}$ with $Var(T_i) < \infty$. Then for $\bar{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$ it holds that

$$\sqrt{n}(\bar{T}_n - \mathbb{E}(T_i)) \xrightarrow{d} N(0, Var(T_i))$$

- Interpretation: sample averages are asymptotically normally distributed.
  $\Rightarrow$ only requires fairly general assumptions (no specific distribution!)
  $\Rightarrow$ can be used to build simple tests
- Only applies to sample averages not any sampled values!

Population/Distribution and Samples



- True distributions of interest but seldomly directly observable
- Sample properties to inform your beliefs of population variables ($\rightarrow$ estimation)
- Be careful what notation is refering to (e.g. $\mathbb{E}$, *Var*)! Convention:
  - Population variables: Greek letters ($\mu, \sigma, \rho, \ldots$)
  - Sample variables: Roman letters ($\bar{x}, s, r, \ldots$)
  - Estimators (for population variables): Hatted greek letters ($\hat{\mu}, \hat{\sigma}, \hat{\rho}, \ldots$)

Recap Chapter 5

Probability space

- Building block of stochastics

Random variables

- Translate from outcome space to the workable reals

Distribution functions

- Density and probability functions characterize underlying distributions
- Allow calculation of expectations, variances, correlations ...

Convergence

- Law of Large Numbers gives consistency of the sample mean
- Central Limit Theorem yields distribution of the sample mean

That's all Folks!

Please take a look at the problem set that we will discuss tomorrow morning.

julian.klix@uni-mannheim.de

L7 3-5, Room 3.43

UNIVERSITÄT MANNHEIM