Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Advanced Quantitative Methods: Mathematics (p)review & statistical estimators

Johan A. Elkink

University College Dublin

26 January 2018

Outline

Course outline

Math (p)review
Matrix algebra
Derivatives

Statistics fundamentals
Expectations and variances
Sampling distributions

Statistical estimators
Finite sample properties
Asymptotic properties

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline

Course outline

Math
(p)review
Matrix algebra
Derivatives

Statistics
fundamentals
Expectations
and variances
Sampling
distributions

Statistical
estimators
Finite sample
properties
Asymptotic
properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Introduction

Topic: advanced quantitative methods in political science.

Or, alternatively: basic econometrics, as applied in political science.

Or, alternatively: linear regression and some non-linear extensions.

Outline

Course outline

Math (p)review
Matrix algebra
Derivatives

Statistics fundamentals
Expectations and variances
Sampling distributions

Statistical estimators
Finite sample properties
Asymptotic properties

Readings

- Peter Kennedy (2008), *A guide to econometrics*. 6th ed., Malden: Blackwell.
- Damodar N. Gujarati (2009), *Basic econometrics*. 5th ed., Boston: McGraw-Hill.
- Julian J. Faraway (2005), *Linear models with R*. Baco Raton: Chapman & Hill.
- Andrew Gelman & Jennifer Hill (2007), *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- William H. Greene (2003), *Econometric analysis*. 5th ed., Upper Saddle River: Prentice Hall.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Topics

1	26/1	Math review & statistical estimators	
2	2/2	Ordinary Least Squares	
3	9/2	Regression diagnostics	
4	16/2	Time-series analysis	
5	23/2	Panel data	*
6	2/3	Causal inference I (instrumental variables)	*
7	9/3	Causal inference II (matching)	
		Study break	
8	6/4	Maximum Likelihood	
9	13/4	Limited dependent variables I	
10	20/4	Limited dependent variables II	
11	27/4	Bootstrap and simulation	

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Frequentist vs Bayesian

Frequentist statistics interprets **probability** as the frequency of occurrance in (hypothetically) many repetitions. E.g. if we throw this dice infinitely many times, what proportion of times would it be heads? We can here also talk of **conditional probabilities**: what would this frequency be if ... and some condition follows.

Bayesian statistics interprets probability as a **belief**: if I throw this dice, what do you think is the chance of getting heads? We can now talk of conditional probabilities in a different way: how would your belief change given that ... and some condition follows.

This course is a course in the frequentist analysis of regression models.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Homeworks

	deadline	*topic*
1	21/2	Linear regression
2	14/3	Time-series & panel data
3	18/4	Causal inference & maximum likelihood
4	2/5	Limited dependent variables & simulation

- 50 % Five homeworks
- 50 % Replication paper
  due May 16, 2018, 5 pm
- No exam

Working with others is a good idea, but write-up needs to be your own.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Grade conversions

Homeworks	UCD	TCD	Homeworks	UCD	TCD
97-100%	A+	A+	54-64%	E+	D
94-96%	A	A	44-53%	E	D
91-93%	A-	A	33-43%	E-	D
88-90%	B+	B+	0-32%	F	F
85-87%	B	B			
83-84%	B-	B			
80-82%	C+	C+			
77-79%	C	C			
74-76%	C	C			
71-73%	D+	C			
68-70%	D	C			
65-67%	D-	C			

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Replication paper

- Find replicable paper *now* and check whether appropriate
- Contact authors asap if you need their data
- See Gary King, "Publication, publication"

It is highly recommended to use the March break for the data analysis of the final assignment, to leave only the write-up to May.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Syllabus and website

- Website: `http://www.joselkink.net/Q2-Spring-2018.php`
- Syllabus downloadable there
- Slides and notes on website
- Data for exercises on website
- Booklet with commands available

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Vectors: examples

$$\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ 1 \\ 5 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 3.23 & 1.30 & 7.89 & 1.00 \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

$\mathbf{v}$ and $\boldsymbol{\beta}$ are **column vectors**, while $\mathbf{w}$ is a **row vector**. When not specified, assume a column vector.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Vectors: transpose

$$\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ 1 \\ 5 \end{bmatrix} \qquad \mathbf{v}' = \begin{bmatrix} 3 & 4 & 1 & 5 \end{bmatrix}$$

Outline

Course outline

Math (p)review
Matrix algebra
Derivatives

Statistics fundamentals
Expectations and variances
Sampling distributions

Statistical estimators
Finite sample properties
Asymptotic properties

Vectors: summation

$$\mathbf{v} = \begin{bmatrix} 3 \\ 5 \\ 9 \\ 1 \\ 3 \end{bmatrix}$$

$$\sum_{i}^{N} v_i = 3 + 5 + 9 + 1 + 3$$

$$= 21$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Vectors: addition

$$\begin{bmatrix} 2 \\ 3 \\ 1 \\ 6 \end{bmatrix} + \begin{bmatrix} 6 \\ 3 \\ 9 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \\ 10 \\ 8 \end{bmatrix}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Vectors: multiplication with scalar

$$\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 6 \end{bmatrix}$$

$$3\mathbf{v} = \begin{bmatrix} 6 \\ 9 \\ 3 \\ 18 \end{bmatrix}$$

Outline

Course outline

Math (p)review
Matrix algebra
Derivatives

Statistics fundamentals
Expectations and variances
Sampling distributions

Statistical estimators
Finite sample properties
Asymptotic properties

Vectors: inner product

$$\mathbf{v} = \begin{bmatrix} 5 \\ 3 \\ 1 \\ 3 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 7 \\ 6 \\ 8 \\ 1 \end{bmatrix}$$

$$\mathbf{v}'\mathbf{w} = 5 \cdot 7 + 3 \cdot 6 + 1 \cdot 8 + 3 \cdot 1 = 64$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Vectors: outer product

$$\mathbf{v} = \begin{bmatrix} 5 \\ 3 \\ 1 \\ 3 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 7 \\ 6 \\ 8 \\ 1 \end{bmatrix}$$

$$\mathbf{v}\mathbf{w}' = \begin{bmatrix} 35 & 30 & 40 & 5 \\ 21 & 18 & 24 & 3 \\ 7 & 6 & 8 & 1 \\ 21 & 18 & 24 & 3 \end{bmatrix}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrices: examples

$$\mathbf{M} = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 2 & 1 \\ 0 & 9 & 8 \end{bmatrix} \qquad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The latter is called an **identity matrix** and is a special type of **diagonal matrix**.

Both are **square matrices**.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrices: transpose

$$\mathbf{X} = \begin{bmatrix} 1 & 3 & 2 \\ 9 & 8 & 8 \\ 9 & 8 & 5 \end{bmatrix}$$

$$\mathbf{X}' = \begin{bmatrix} 1 & 9 & 9 \\ 3 & 8 & 8 \\ 2 & 8 & 5 \end{bmatrix}$$

$$(\mathbf{A}')' = \mathbf{A}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrices: indexing

$$\mathbf{X}_{4\times 3} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix}$$

Always rows first, columns second.

So $\mathbf{X}_{n\times k}$ is a matrix with n rows and k columns. $\mathbf{y}_{n\times 1}$ is a column vector with n elements.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrices: addition

$$\begin{bmatrix} 4 & 2 & 1 & 3 \\ 6 & 5 & 3 & 2 \\ 1\frac{1}{2} & 4 & 2\frac{1}{3} & 0 \end{bmatrix} + \begin{bmatrix} 7 & 8 & 2 & 0 \\ -3 & -2 & 3 & 1\frac{1}{2} \\ 4 & 2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 10 & 3 & 3 \\ 3 & 3 & 6 & 3\frac{1}{2} \\ 5\frac{1}{2} & 6 & 3\frac{1}{3} & 1 \end{bmatrix}$$

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrices: multiplication with scalar

$$\mathbf{X} = \begin{bmatrix} 4 & 2 & 1 & 3 \\ 6 & 5 & 3 & 2 \\ 1\frac{1}{2} & 4 & 2\frac{1}{3} & 0 \end{bmatrix}$$

$$4\mathbf{X} = \begin{bmatrix} 16 & 8 & 4 & 12 \\ 24 & 20 & 12 & 8 \\ 6 & 16 & 7\frac{1}{3} & 0 \end{bmatrix}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrices: multiplication

$$\mathbf{A} = \begin{bmatrix} 6 & 4 \\ 1 & 3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 3 & 2 & 3 \\ 4 & 2 & 1 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 6\cdot 3 + 4\cdot 4 & 6\cdot 2 + 4\cdot 2 & 6\cdot 3 + 4\cdot 1 \\ 1\cdot 3 + 3\cdot 4 & 1\cdot 2 + 3\cdot 2 & 1\cdot 3 + 3\cdot 1 \end{bmatrix} = \begin{bmatrix} 34 & 20 & 22 \\ 15 & 8 & 6 \end{bmatrix}$$

$$(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$$

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

$$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Special matrices

Symmetric matrix	$\mathbf{A}' = \mathbf{A}$
Idempotent matrix	$\mathbf{A}^2 = \mathbf{A}$
Positive-definite matrix	$\mathbf{x}'\mathbf{A}\mathbf{x} > 0 \quad \forall \quad \mathbf{x} \neq 0$
Positive-semidefinite matrix	$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \quad \forall \quad \mathbf{x} \neq 0$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrix rank

The **rank** of a matrix is the maximum number of independent columns or rows in the matrix. Columns of a matrix $\mathbf{X}$ are independent if for any $\mathbf{v} \neq 0$, $\mathbf{Xv} \neq 0$

$$r(\mathbf{A}) = r(\mathbf{A}') = r(\mathbf{A}'\mathbf{A}) = r(\mathbf{A}\mathbf{A}')$$
$$r(\mathbf{AB}) = min(r(\mathbf{A}), r(\mathbf{B}))$$

Matrix rank: example 1

$$\mathbf{A} = \begin{bmatrix} 3 & 5 & 1 \\ 2 & 2 & 1 \\ 1 & 4 & 2 \end{bmatrix}$$

For matrix $\mathbf{A}$, the three columns are independent and $r(\mathbf{A}) = 3$. There is no $\mathbf{v} \neq 0$ such that $\mathbf{A}\mathbf{v} = 0$ (of course, if $\mathbf{v} = 0$, $\mathbf{A}\mathbf{v} = \mathbf{A}0 = 0$).

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrix rank: example 2

$$\mathbf{B} = \begin{bmatrix} 3 & 5 & 9 \\ 2 & 2 & 6 \\ 1 & 4 & 3 \end{bmatrix}$$

For matrix $\mathbf{B}$ the first and the last column vectors are linear combinations of each other: $\mathbf{b}_{\bullet 1} = \frac{1}{3}\mathbf{b}_{\bullet 3}$. At most two columns ($\mathbf{b}_{\bullet 1}$ and $\mathbf{b}_{\bullet 2}$ or $\mathbf{b}_{\bullet 2}$ and $\mathbf{b}_{\bullet 3}$) are independent, so $r(\mathbf{B}) = 2$. We could construct a matrix $\mathbf{v}$ such that $\mathbf{Bv} = 0$, namely $\mathbf{v} = \begin{bmatrix} 1 & 0 & -\frac{1}{3} \end{bmatrix}'$ (or any $\mathbf{v} = \begin{bmatrix} \alpha & 0 & -\frac{1}{3}\alpha \end{bmatrix}'$).

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrix rank: example 3

$$\mathbf{C} = \begin{bmatrix} 3 & 5 & 11\frac{1}{2} \\ 2 & 2 & 7 \\ 1 & 4 & 5 \end{bmatrix}$$

$r(\mathbf{C}) = 2$. In this case, one cannot express one of the column vectors as a linear combination of another column vector, but one can express any of the column vectors as a linear combination of the two other column vectors. For example, $\mathbf{c}_{\bullet 3} = 3\mathbf{c}_{\bullet 1} + \frac{1}{2}\mathbf{c}_{\bullet 2}$ and thus when $\mathbf{v} = \begin{bmatrix} 3 & \frac{1}{2} & -1 \end{bmatrix}'$, $\mathbf{C}\mathbf{v} = 0$.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrix inverse

The **inverse** of a matrix is the matrix one would have to multiply with to get the identity matrix, i.e.:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

$$(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$$
$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Matrices: trace

The **trace** of a matrix is the sum of the diagonal elements.

$$\mathbf{A} = \begin{bmatrix} 4 & 3 & 1 & 8 \\ 2 & 8 & 5 & 5 \\ 6 & 7 & 3 & 4 \\ 1 & 2 & 0 & 1 \end{bmatrix}$$

$$tr(\mathbf{A}) = sum(diag(\mathbf{A})) = 4 + 8 + 3 + 1 = 16$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

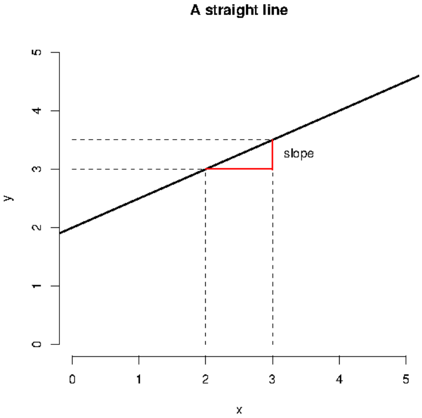
4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

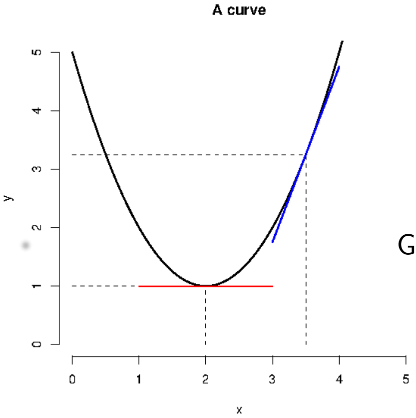
Derivatives of straight lines



$$y = \frac{1}{2}x + 2$$
$$\frac{dy}{dx} = \frac{1}{2}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Derivatives of curves



$$y = x^2 - 4x + 5$$

$$\frac{dy}{dx} = 2x - 4$$

Generally:

$$\frac{d(ax^b)}{dx} = bax^{b-1}$$

$$\frac{d(a+b)}{dx} = \frac{da}{dx} + \frac{db}{dx}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Finding location of peaks and valleys

A maximum or minimum value of a curve can be found by setting the derivative equal to zero.

$$y = x^2 - 4x + 5$$
$$\frac{dy}{dx} = 2x - 4 = 0$$
$$2x = 4$$
$$x = \frac{4}{2} = 2$$

So at $x = 2$ we will find a (local) maximum or minimum value.

You would need a second derivative to know whether it is a minimum or a maximum (or look at the plot). In this case $\frac{d(2x-4)}{dx} = 2$, which is positive, so it is a minimum.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Derivatives in regression

This concept of finding the minimum or maximum is crucial for regression analysis.

- With **ordinary least squares** (OLS), we estimate the β-coefficients by finding the minimum of the sum of squared errors.
- With **maximum likelihood** (ML), we estimate the β-coefficients by finding the maximum point of the (log)likelihood function.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Derivatives of matrices

The derivative of a matrix consists of the matrix containing the derivatives of each element of the original matrix.

$$\mathbf{M} = \begin{bmatrix} x^2 & 2x & 3 \\ 2x+4 & 3x^3 & 2x \\ 3 & 2 & 4x \end{bmatrix}$$

$$\frac{d\mathbf{M}}{dx} = \begin{bmatrix} 2x & 2 & 0 \\ 2 & 9x^2 & 2 \\ 0 & 0 & 4 \end{bmatrix}$$

Outline

Course outline

Math
(p)review
Matrix algebra
Derivatives

Statistics
fundamentals
Expectations
and variances
Sampling
distributions

Statistical
estimators
Finite sample
properties
Asymptotic
properties

Derivatives of matrices

$$\frac{\partial(\mathbf{v}'\mathbf{x})}{\partial\mathbf{x}} = \mathbf{v}$$

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial\mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$$

$$\frac{\partial(\mathbf{B}\mathbf{x})}{\partial\mathbf{x}} = \mathbf{B}'$$

$$\frac{\partial^2(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial\mathbf{x}d\mathbf{x}'} = \mathbf{A} + \mathbf{A}'$$

(Harville, 2008, 289–327)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Logarithms

Some rules about logarithms, which are important for understanding ML:

$$
\begin{aligned}
\log ab &= \log a + \log b \\
\log e^a &= a \\
\log a^b &= b \log a \\
\log a > \log b \quad &\text{iff} \quad a > b \\
\frac{d(\log a)}{da} &= \frac{1}{a}
\end{aligned}
$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Expected value

The **expected value** of a discrete random variable x is:

$$E(x) = \sum_{i}^{N} P(x_i) \cdot x_i,$$

whereby N is the total number of possible outcomes in S.

For a continuous variable, the equivalent is:

$$E(x) = \int x f(x) dx,$$

with $f(x)$ the probability density function of x.

The expected value is the **mean** (μ) of a random variable (not to be confused with the mean of a particular set of observations).

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Moments

- moments are $E(x)^n, \; n = 1, 2, ...$
  - So μ_x is the first moment of x
- central moments are $E(x - E(x))^n, \; n = 1, 2, ...$
- absolute moments are $E(|x|)^n, \; n = 1, 2, ...$
- absolute central moments are $E(|x - E(x)|)^n, \; n = 1, 2, ...$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Variance

The second central moment is called the **variance**, so:

$$var(x) = E(x - E(x))^2 = E(x - \mu_x)^2$$
$$= \sum_i^N P(x_i) \cdot (x_i - E(x))^2$$

The **standard deviation** is the square root of the variance:

$$sd(x) = \sqrt{var(x)} = \sqrt{E(x - E(x))^2}$$

For a continuous variable, the population variance is:

$$var(x) = \int x^2 f(x) dx - \mu^2$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Covariance

The **covariance** of two random variables x and y is defined as:

$$cov(x, y) = E[(x - E(x))(y - E(y))]$$

If x and y are **independent** of each other, $cov(x, y) = 0$.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Variance of a sum

$$var(x+y) = var(x) + var(y) + 2cov(x,y)$$
$$var(x-y) = var(x) + var(y) - 2cov(x,y)$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Probabilities: definition

Frequentist definition:

$$P(x) = lim_{n\rightarrow\infty}\frac{f(x)}{n},$$

where P is the probability, n the number of trials and f the frequency of event x occurring.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Probabilities: properties

For event E in **sample space** S:

$$
\begin{aligned}
0 \leq P(E) &\leq 1 \\
P(S) &= 1 \\
P(\neg E) &= 1 - P(E) \\
P(E_1 \cup E_2) &= P(E_1) + P(E_2) \quad \text{if } E_1 \text{ and } E_2 \text{ are mutually exclusive}
\end{aligned}
$$

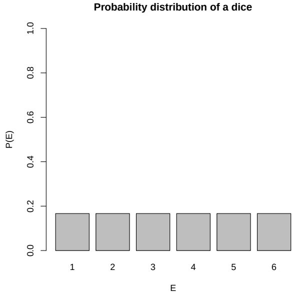
A sample space is the set of all possible outcomes, while an event is a specific subset of the sample space.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Probability distribution

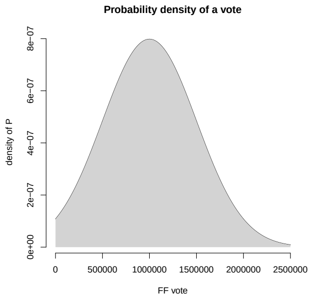
A probability distribution of a discrete variable presents the probabilities for each possible outcome.

Because $P(S) = 1$, the bars add up to 1.



Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Probability distribution



A probability distribution of a continuous variable presents the probability density for each possible outcome. The surface under the plot represents the probability of the outcome being within a particular range.

Because $P(S) = 1$, the surface under the entire plot is 1.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

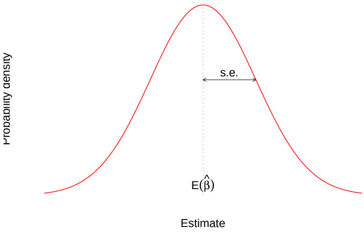
Sampling distribution

Imagine, that instead of having one sample, we take many samples.

If we do the same estimation in each of those randomly selected samples, we would get different results each time.

The distribution of these different estimates is the **sampling distribution** of the estimate.

Sampling distribution



The sampling distribution is a probability density function, with as mean the expected value of $\hat{\beta}$.

The spread of this distribution is indicated by the **standard error** (s.e.).

$$se_{\hat{\beta}} = \sqrt{var(\hat{\beta})}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Estimation

"Problems of estimation are those in which it is required to estimate the value of one or more of the population parameters from a random sample of the population."

(Fisher, 1922, 310)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Estimators

- Ordinary Least Squares (OLS)
- Generalized Least Squares (GLS)
- Maximum Likelihood (ML)
- Simulated Maximum Likelihood (SML)
- General Method of Moments (GMM)
- Bayesian (usually Markov Chain Monte Carlo) (MCMC)
- etc.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Estimator criteria

Given that there are often many possible estimators for a particular population value, we want to be able to **evaluate** which one is most appropriate.

We can make a distinction between two types of criteria:

- **Finite sample** properties - how well does the estimator do given a limited sample size?
- **Asymptotic** properties - how well does the estimator do as the sample size gets infinitely large?

(We will assume single parameter estimations (β), rather than multivariate ones ($\boldsymbol{\beta}$) for the remainder of these slides.)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Unbiasedness

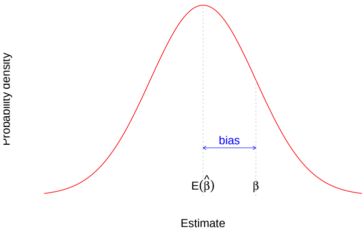
The **bias** of an estimator is the difference between the expected value of the sample distribution and the true value of the parameter to be estimated:

$$bias_{\hat{\beta}} = E(\hat{\beta}) - \beta$$

So an **unbiased estimator** is an estimator where $E(\hat{\beta}) = \beta$.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Bias: sampling distribution



For a biased estimator, $E(\hat{\beta}) \neq \beta$, and the bias is $E(\hat{\beta}) - \beta$.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Example: variance estimation

It can be shown (see notes) that, if s^2 is the **sample variance** and σ^2 the **population variance**, that:

$$E(s^2) = \frac{n-1}{n}\sigma^2,$$

in other words, s^2 is a **biased estimator** of σ^2.

$$E(\frac{n}{n-1}s^2) = \frac{n}{n-1}E(s^2) = \frac{n}{n-1} \cdot \frac{n-1}{n}\sigma^2 = \sigma^2,$$

so when we estimate the population variance, we calculate $\frac{1}{n-1}\sum_i^n (x_i - \bar{x})^2$ instead of $\frac{1}{n}\sum_i^n (x_i - \bar{x})^2$.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Unbiasedness: vector of coefficients

If $\boldsymbol{\beta}$ is a vector of coefficients, rather than a single parameter β, it simply still holds that an unbiased estimator is one where $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ and a biased estimator where it is not.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

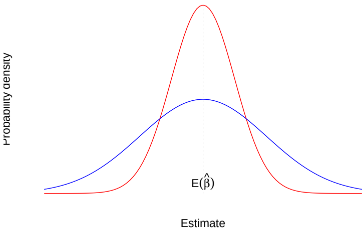
Efficiency

The estimator whose sampling distribution has the lowest variance is the more efficient estimator.

The most efficient unbiased estimator is called the **best unbiased** estimator.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Efficiency: sampling distribution



$E(\hat{\beta})$ is the same for both estimators, but the estimator with the red sampling distribution is more efficient than the one with the blue sampling distribution.

$$se_{\hat{\beta}_{blue}} > se_{\hat{\beta}_{red}}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

BLUE

In the context of linear models, we often talk of the **best linear unbiased estimator** (BLUE), which is the estimator which is linear, unbiased, and has the lowest sampling variance of all possible unbiased linear estimators.

When the assumptions underlying OLS hold, OLS is BLUE.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Efficiency: vector of coefficients

Often, $\boldsymbol{\beta}$ is a vector instead of just one parameter β. Instead of just the variance of β, we have a **variance-covariance matrix**:

$$var(\boldsymbol{\beta}) = \begin{bmatrix} var(\beta_1) & cov(\beta_1, \beta_2) & \dots & cov(\beta_1, \beta_k) \\ cov(\beta_2, \beta_1) & var(\beta_2) & \dots & cov(\beta_2, \beta_k) \\ \dots & \dots & \dots & \dots \\ cov(\beta_k, \beta_1) & cov(\beta_k, \beta_2) & \dots & var(\beta_k) \end{bmatrix}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Efficiency: vector of coefficients

In this case we can consider various criteria for efficiency:

- smallest trace
- smallest determinant
- smallest variance of any linear combination of its elements
- some weighted sum of the variances and covariances

$$E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{W}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]$$

If **W** is selected such that this equation cannot lead to a negative result, minimizing this expectation leads to the most efficient estimator on all the above grounds.

(Kennedy, 2008, 34)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Mean Square Error

Sometimes an estimator can be slightly biased, but be so much more efficient, that is becomes preferable.

So an estimator which is BLUE is not necessarily the most efficient estimator - it is simply the most efficient *unbiased* estimator.

The **Mean Square Error** (MSE) is the sum of the variance and the square of the bias of an estimator:

$$\begin{aligned} MSE_{\hat{\beta}} &= E[(\hat{\beta} - \beta)^2] \\ &= var(\hat{\beta}) + E(\hat{\beta} - \beta)^2 \end{aligned}$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

MSE: vector of coefficients

When $\boldsymbol{\beta}$ is a vector of coefficients instead of a single parameter β, we could look at the **MSE matrix**:

$$\textbf{MSE} = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'$$

Or we could just look at the trace of this matrix.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Outline

1 Course outline

2 Math (p)review

Matrix algebra

Derivatives

3 Statistics fundamentals

Expectations and variances

Sampling distributions

4 Statistical estimators

Finite sample properties

Asymptotic properties

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Asymptotics

The **asymptotic properties** of an estimator concern the estimator's sampling distribution in extremely (or infinitely) large samples.

We need these properties when we cannot provide the same proofs of unbiasedness, efficiency, etc. for small samples. To what extend the properties hold for small samples is then left for further exploration (e.g. through simulation).

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Limits

"The real-valued sequence $\{x_n\}$ has the real number x^* for its **limit**, or converges to x^*, if for any positive ε, no matter how small, it is possible to find a positive integer N such that for all integers n greater than N, $|x_n - x^*| < \varepsilon$".

We can write:

$$\lim_{n \to \infty} x_n = x^*.$$

(Davidson and MacKinnon, 1993, 102)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Limits: examples

$$\lim_{n \to \infty} \frac{1}{n} = 0$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Limits: examples

$$\lim_{n\to\infty} \frac{1}{n} = 0$$

$$\lim_{n\to\infty} \frac{2n+1}{n-1} = 2$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Limits: examples

$$\lim_{n\to\infty} \frac{1}{n} = 0$$

$$\lim_{n\to\infty} \frac{2n+1}{n-1} = 2$$

$$\lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Probability limits

$$P(\|\mathbf{x}_n - \mathbf{x}^*\| > \varepsilon) < \delta,$$

whereby ε and δ are arbitrarily small, positive real numbers. We can write:

$$\operatorname*{plim}_{n\to\infty} \mathbf{x}_n = \mathbf{x}^*.$$

(Davidson and MacKinnon, 1993, 103)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Convergence in distribution

$$\lim_{n\to\infty} P(\mathbf{x}_n \leq \mathbf{b}) = P(\mathbf{x}^* \leq \mathbf{b})$$

for any arbitrary $\mathbf{b}$. In other words, the probability distribution of $\{\mathbf{x}_n\}$ approaches that of $\mathbf{x}^*$ as n increases. This can be written as:

$$\mathbf{x}_n \xrightarrow{D} \mathbf{x}^*.$$

(Davidson and MacKinnon, 1993, 107)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Asymptotic unbiasedness

Just was we can look at the bias in finite samples, we can also talk of the **asymptotic bias** in infinitely large samples:

$$asy.bias = \lim_{n \to \infty} (E(\hat{\beta}) - \beta)$$

(but see Greene, 2003, 917)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Asymptotic unbiasedness: example

$$E(s^2) = \frac{n-1}{n}\sigma^2,$$

so s^2 is a **biased** estimator of σ^2.

$$\lim_{n\to\infty} E(s^2) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2,$$

so s^2 is an **asymptotically unbiased** estimator of σ^2.

(see for the asymptotic variance of this estimator Greene, 2003, 917–918)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Consistency

"A statistic satisfies the criterion of consistency, if, when it is calculated from the whole population, it is equal to the required parameter."

(Fisher, 1922, 309)

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Consistency

An estimator is **consistent** iff:

$$\operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$$

In other words, the **asymptotic variance** of $\hat{\boldsymbol{\beta}}$ becomes very small and $\hat{\boldsymbol{\beta}}$ is **asymptotically unbiased**, so that the probability distribution of $\hat{\boldsymbol{\beta}}$ "collapses" to a very tight distribution around $\boldsymbol{\beta}$.

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Consistency: sample mean

The **Central Limit Theorem** states:

$$E(\bar{x}) = \mu_x \qquad var(\bar{x}) = \frac{\sigma^2}{n}$$

therefore:

$$\operatorname*{plim}_{n\to\infty} E(\bar{x}) = \mu_x \qquad \operatorname*{plim}_{n\to\infty} var(\bar{x}) = 0$$

The mean of a random sample ($\bar{x}$) is a **consistent estimator** of the mean of the population (μ_x).

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Efficiency

"The criterion of efficiency is satisfied by those statistics which, when derived from large samples, tend to a normal distribution with the least possible standard deviation."

(Fisher, 1922, 310)

Outline

Course outline

Math (p)review
Matrix algebra
Derivatives

Statistics fundamentals
Expectations and variances
Sampling distributions

Statistical estimators
Finite sample properties
Asymptotic properties

Asymptotic variance

The **asymptotic variance** of an estimator is:

$$\lim_{n \to \infty} var(\hat{\beta})$$

Outline
Course outline
Math (p)review
Matrix algebra
Derivatives
Statistics fundamentals
Expectations and variances
Sampling distributions
Statistical estimators
Finite sample properties
Asymptotic properties

Asymptotic efficiency

"An estimator is **asymptotically efficient** if it is consistent, asymptotically normally distributed, and has an asymptotic covariance matrix that is not larger than the asymptotic covariance matrix of any other consistent, asymptotically normally distributed estimator."

(Greene, 2003, 71)

Appendix

Least squares

We can look at the difference between the predicted and the observed values of the dependent variable:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$$

and we want this prediction error to be as small as possible.

Many options:

- minimizing absolute errors, $min(|\mathbf{e}|)$
- minimizing squared errors, $min(\mathbf{e}'\mathbf{e})$
- using weights, $min(\mathbf{e}'\mathbf{W}\mathbf{e})$
- etc.

OLS minimizes $\mathbf{e}'\mathbf{e}$.

R^2

R^2 represents the explained variance in $\mathbf{y}$, as a proportion of the total variance in $\mathbf{y}$.

R^2

Only appropriate when:

- using OLS for estimation
- explained variation is linear only
- there is an intercept in the model

R^2

R^2 represents the explained variance in $\mathbf{y}$, as a proportion of the total variance in $\mathbf{y}$.

- OLS maximizes R^2 implicitly

R^2

R^2 represents the explained variance in $\mathbf{y}$, as a proportion of the total variance in $\mathbf{y}$.

- OLS maximizes R^2 implicitly
- High R^2 can be caused by:
    - Using a lot of regressors
    - Including lagged dependent variable

R^2

R^2 represents the explained variance in $\mathbf{y}$, as a proportion of the total variance in $\mathbf{y}$.

- OLS maximizes R^2 implicitly
- High R^2 can be caused by:
  - Using a lot of regressors
  - Including lagged dependent variable
- No good equivalent for non-linear of models

R^2

R^2 represents the explained variance in $\mathbf{y}$, as a proportion of the total variance in $\mathbf{y}$.

- OLS maximizes R^2 implicitly
- High R^2 can be caused by:
  - Using a lot of regressors
  - Including lagged dependent variable
- No good equivalent for non-linear of models
- When interested in causal effect, high R^2 is not all too interesting

R^2

“These measures of goodness of fit have a fatal attraction. Although it is generally conceded among insiders that they do not mean a thing, high values are still a source of pride and satisfaction to their authors, however hard they may try to conceal these feelings.”

(Cramer 1987, 253, 45-51, as cited in Kennedy 2008, 27)

R^2 and least squares

Too much emphasis on these criteria can lead to **overfitting**, where you get excellent results for the sample at hand, but not if you would use the same estimates on any other data.

One way to reduce the problem of overfitting is to **split the sample** in two, use one half for estimation, and then use the estimated values to predict the dependent variable in the other half of the sample, and check errors and R^2.

Likelihood

"The likelihood that any parameter (or set of parameters) should have any assigned value (or set of values) is proportional to the probability that if this were so, the totality of observations should be that observed."

(Fisher, 1922, 310)

References

Maximum Likelihood

The likelihood is proportional to the probability of observing the data you have (give or take some arbitrarily small deviation), given some parameter estimate:

$$L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \alpha P(\mathbf{y}, \mathbf{X}|\boldsymbol{\beta})$$

The **likelihood function** is thus proportional to the **probability density function** of the given sample, as a function of the parameter values.

The estimator that maximizes this function also maximizes this probability density function and is the **Maximum Likelihood Estimator** (MLE or ML).

(Fisher, 1922)

Sufficiency

"A statistic satisfies the criterion of sufficiency when no other statistic which can be calculated from the same sample provides any additional information as to the value of the parameter to be estimated."

(Fisher, 1922, 310)

Computational costs

Used to be a major problem ...

... not so much anymore.

Still worth considering for very large datasets.

E.g. $|(\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}|$ in a spatial regression model.

math
(p)review

References

Cramer, Jan Solomon. 1987. "Mean and variance of R^2 in small and moderate samples." *Journal of econometrics* 35(2):253–266.

Davidson, Russell and James G. MacKinnon. 1993. *Estimation and inference in econometrics*. Oxford: Oxford University Press.

Faraway, Julian J. 2005. *Linear models with R*. Boca Raton: Chapman & Hall.

Fisher, Ronald A. 1922. "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 222(594–604):309–368.
URL: *http://rsta.royalsocietypublishing.org/content/222/594-604/309*

Gelman, Andrew and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Analytical Methods for Social Research Cambridge: Cambridge University Press.

Gentle, James E. 2007. *Matrix algebra: theory, computations, and applications in statistics*. New York: Springer.

Greene, William H. 2003. *Econometric analysis*. 5th ed. Upper Saddle River: Prentice Hall.

Gujarati, Damodar N. 2009. *Basic econometrics*. 5th ed. Boston: McGraw-Hill.

Harville, David A. 2008. *Matrix algebra from a statistician's perspective*. New York: Springer-Verlag.

Kennedy, Peter. 2008. *A guide to econometrics*. 6th ed. Malden, MA: Blackwell.