# Advanced Quantitative Methods: Mathematics (p)review

Jos Elkink

University College Dublin

January 19, 2011

# Outline

# Introduction

Topic: advanced quantitative methods in political science.

# Introduction

Topic: advanced quantitative methods in political science.

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

# 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.

# 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.

# 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.

# Topics

| | | |
|---|---|---|
| 1 | 19/1 | Mathematics review |
| 2 | 26/1 | Statistical estimators |
| 3 | 2/2 | Ordinary Least Squares |
| 4 | 9/2 | Hypothesis testing |
| 5 | 16/2 | Specification, multicollinearity and heteroscedasticity |
| 6 | 23/2 | Autocorrelation |
| 7 | 2/3 | Time-series analysis |
| | | *Study break* |
| 8 | 23/3 | Maximum Likelihood |
| 9 | 30/3 | Limited dependent variables |
| 10 | 6/4 | Bootstrap and simulation |
| 11 | 13/4 | Multilevel data |
| 12 | 20/4 | Panel data |

# 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.

# 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.

# 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.

# Homeworks

- 50 % Weekly homeworks, top ten grades
  due before class the following week
- 50 % Replication paper
  due April 29, 2011, 5 pm
- No exam

# Homeworks

- 50 % Weekly homeworks, top ten grades
  due before class the following week
- 50 % Replication paper
  due April 29, 2011, 5 pm
- No exam

*Working with others is a good idea*

# Grade conversions

| Homeworks | UCD | TCD | Homeworks | UCD | TCD |
|---|---|---|---|---|---|
| 95-100% | A+ | A+ | 45-49% | E+ | D |
| 91-94% | A | A+ | 37-44% | E | D |
| 87-90% | A- | A | 33-36% | E- | D |
| 83-86% | B+ | A | 0-32% | F | F |
| 79-82% | B | B+ | | | |
| 75-78% | B- | B+ | | | |
| 70-74% | C+ | B | | | |
| 66-69% | C | B | | | |
| 62-65% | C- | C+ | | | |
| 58-61% | D+ | C+ | | | |
| 54-57% | D | C | | | |
| 50-53% | D- | C | | | |

# Replication paper

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

# Syllabus and website

- Website: http://www.joselkink.net/teaching
- Syllabus downloadable there
- Slides and notes on website
- Data for exercises on website

# Outline

1. Course outline
2. Matrix algebra
3. Derivatives
4. Probabilities and probability distributions
5. Expectations and variances

# Vectors: examples

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

# Vectors: examples

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

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

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

# Vectors: transpose

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

# Vectors: summation

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

$$\sum_{i}^{N} v_i = 3 + 5 + 9 + 1 + 3$$
$$= 21$$

# Vectors: addition

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

# Vectors: multiplication with scalar

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

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

# Vectors: inner product

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

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

# Vectors: outer product

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

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

# Matrices: examples

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

# 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**.

# Matrices: transpose

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

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

# Matrices: transpose

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

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

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

# Matrices: indexing

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

# 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.

# 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.

# Matrices: addition

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

# Matrices: addition

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

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

# Matrices: multiplication with scalar

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

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

# Matrices: multiplication

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

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

# Matrices: multiplication

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

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

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

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

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

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

# 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 | $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \quad \forall \quad \mathbf{x} \neq 0$ |

# 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 $\mathbf{v} \neq 0$, $\mathbf{X}\mathbf{v} \neq 0$

# 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 $\mathbf{Av} = 0$ (of course, if $\mathbf{v} = 0$, $\mathbf{Av} = \mathbf{A}0 = 0$).

# 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 $\mathbf{v} = \begin{bmatrix} \alpha & 0 & -\frac{1}{3}\alpha \end{bmatrix}'$).

# 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 $\mathbf{v} = \begin{bmatrix} 3 & \frac{1}{2} & -1 \end{bmatrix}'$, $\mathbf{Cv} = 0$.

# Matrix inverse

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

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

# Matrix inverse

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

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

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

# Matrices: trace

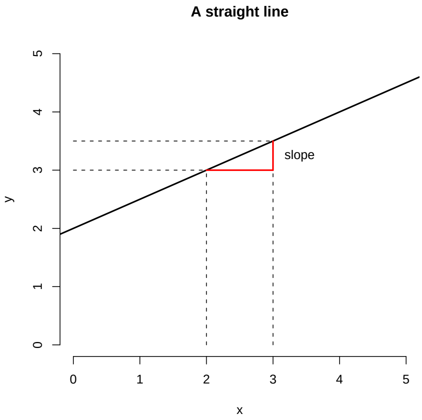
The **trace** of a matrix is the sum of the diagonal elements.

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

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

# Outline

# Derivatives of straight lines

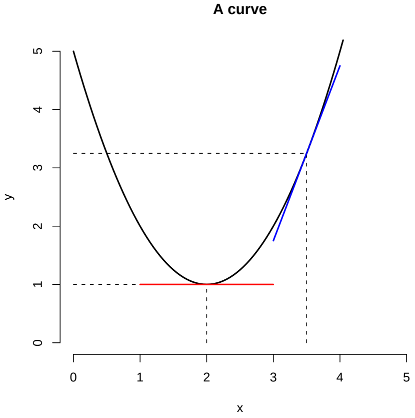



$$y = \frac{1}{2}x + 2$$

$$\frac{dy}{dx} = \frac{1}{2}$$

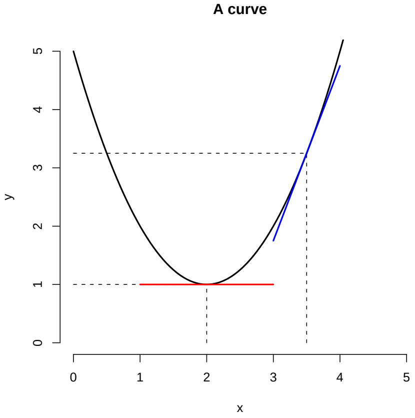
# Derivatives of curves

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



# Derivatives of curves

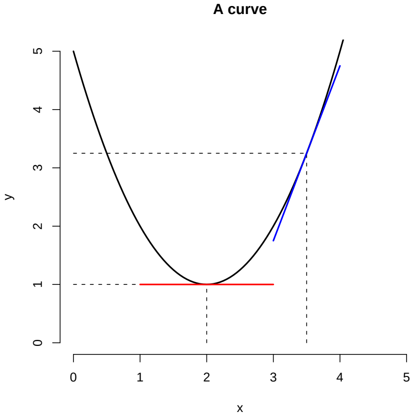
$$y = x^2 - 4x + 5$$




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

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

# Derivatives of curves




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

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

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

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

# Finding location of peaks and valleys

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

# Finding location of peaks and valleys

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

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

$$\frac{dy}{dx} = 2x - 4 = 0$$

$$2x = 4$$

$$x = \frac{4}{2} = 2$$

# 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 - the plot shows that in this case it is a minimum.

# Derivatives in regression

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

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

# Derivatives of matrices

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

# Derivatives of matrices

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

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

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

# Derivatives of matrices

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

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

$$\frac{d(\mathbf{B}\mathbf{x})}{d\mathbf{x}} = \mathbf{B}$$

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

# Outline

# Probabilities: definition

Frequentist definition:

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

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

# Probabilities: properties

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

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

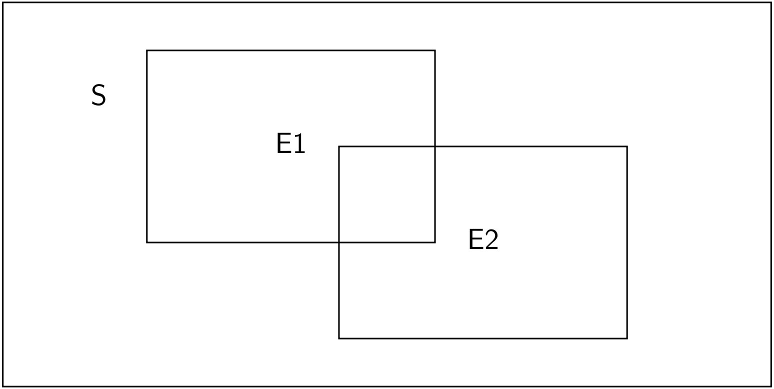
A sample space is the set of all possible outcomes, while an event is a specific subset of the sample space.

# Conditional probability

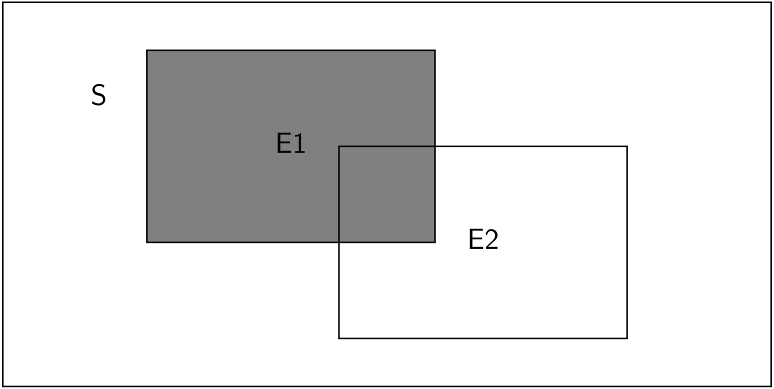
$P(E_1|E_2)$ denotes the probability of event $E_1$ happening, given that event $E_2$ occurred.

If $P(E_1|E_2) = P(E_1)$ then $E_1$ and $E_2$ are independent events.

# Venn diagram

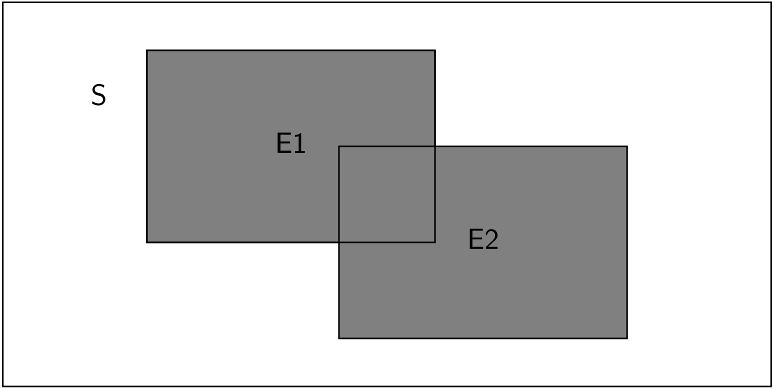


# Venn diagram



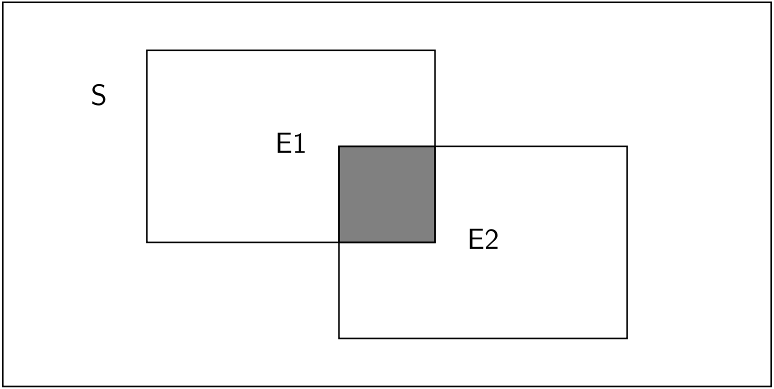

$$P(E_1)$$

# Venn diagram



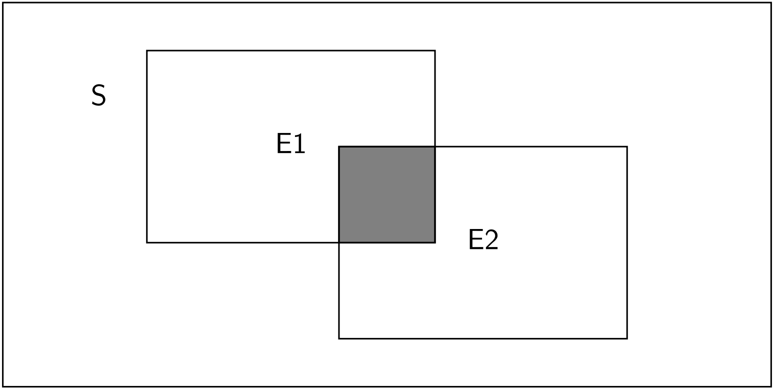

$$P(E_1 \cup E_2)$$

# Venn diagram




$$P(E_1 \cap E_2)$$

# Conditional probability




$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

# Bayes' theorem

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$
$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

therefore:

$$P(E_1 \cap E_2) = P(E_1|E_2) \cdot P(E_2)$$
$$P(E_1 \cap E_2) = P(E_2|E_1) \cdot P(E_1)$$

# Bayes' theorem

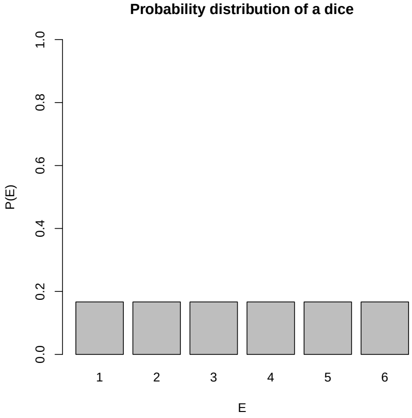
$$
\begin{aligned}
P(E_1|E_2) &= \frac{P(E_1 \cap E_2)}{P(E_2)} \\
&= \frac{P(E_2|E_1) \cdot P(E_1)}{P(E_2)} \\
&= \frac{P(E_2|E_1) \cdot P(E_1)}{P(E_2|E_1) \cdot P(E_1) + P(E_2|\neg E_1) \cdot P(\neg E_1)}
\end{aligned}
$$

# Bayes' theorem

$$
\begin{aligned}
P(E_1|E_2) &= \frac{P(E_1 \cap E_2)}{P(E_2)} \\
&= \frac{P(E_2|E_1) \cdot P(E_1)}{P(E_2)} \\
&= \frac{P(E_2|E_1) \cdot P(E_1)}{P(E_2|E_1) \cdot P(E_1) + P(E_2|\neg E_1) \cdot P(\neg E_1)}
\end{aligned}
$$

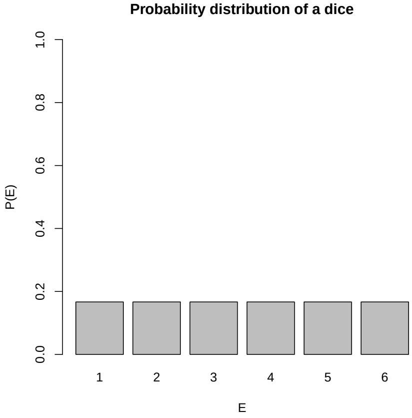
Note that while Bayesian inference is based on Bayes' theorem, the theorem itself holds regardless of your definition of probability.

# Probability distribution




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

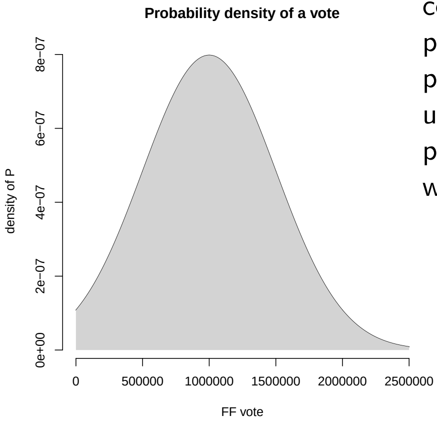
# Probability distribution




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

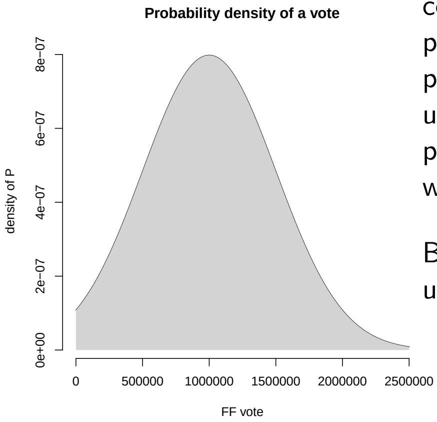
Because $P(S) = 1$, the bars add up to 1.

# 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.

# 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

# Expected value

The **expected value** of a 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 $S$.

# Expected value

The **expected value** of a 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 $S$.

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

# Moments

- moments are $E(x)^n$, $n = 1, 2, ...$
  - So $\mu_x$ 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 $E(|x - E(x)|)^n$, $n = 1, 2, ...$

# Variance

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

$$
\begin{aligned}
var(x) &= E(x - E(x))^2 = E(x - \mu_x)^2 \\
&= \sum_i^N P(x_i) \cdot (x_i - E(x))^2
\end{aligned}
$$

# 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:

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

# Covariance

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

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

# 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, $cov(x, y) = 0$.

# Variance of a sum

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