ECS764P: LECTURE 1

Introduction to Probabilities

Dr Fredrik Dahlqvist

ABOUT ECS764P

Queen Mary
University of London

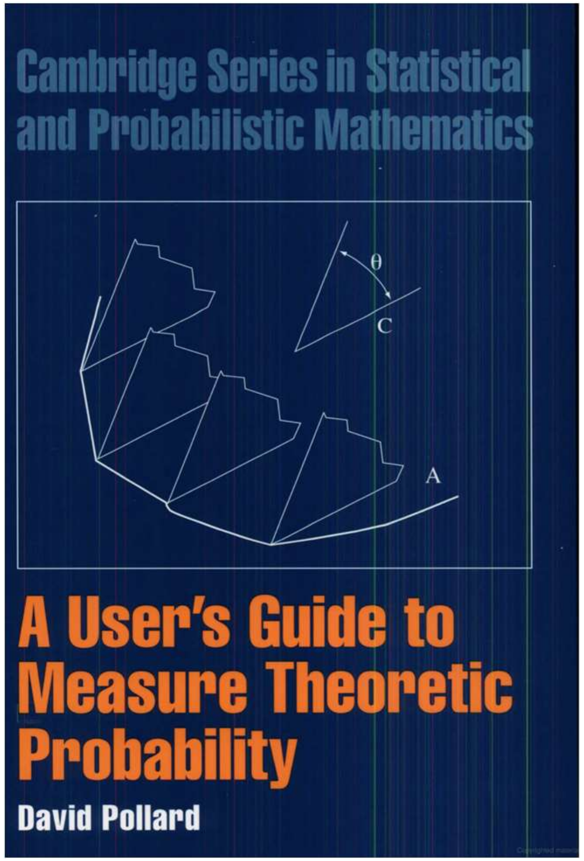
- **Lectures**: Wednesdays 13:00 to 15:00.
- **4 Labs:** 12:00 to 14:00 (TB Lab) on:
  - Thursday 10 October
  - Thursday 31 October
  - Thursday 14 November
  - Thursday 5 December
- **4 pieces of coursework:** The “Applied” part of “Applied Statistics”. 10% of the marks each, submit Jupyter notebook, code must run without bugs!
- **Exam** (60% of the marks)

MODULE CONTENTS

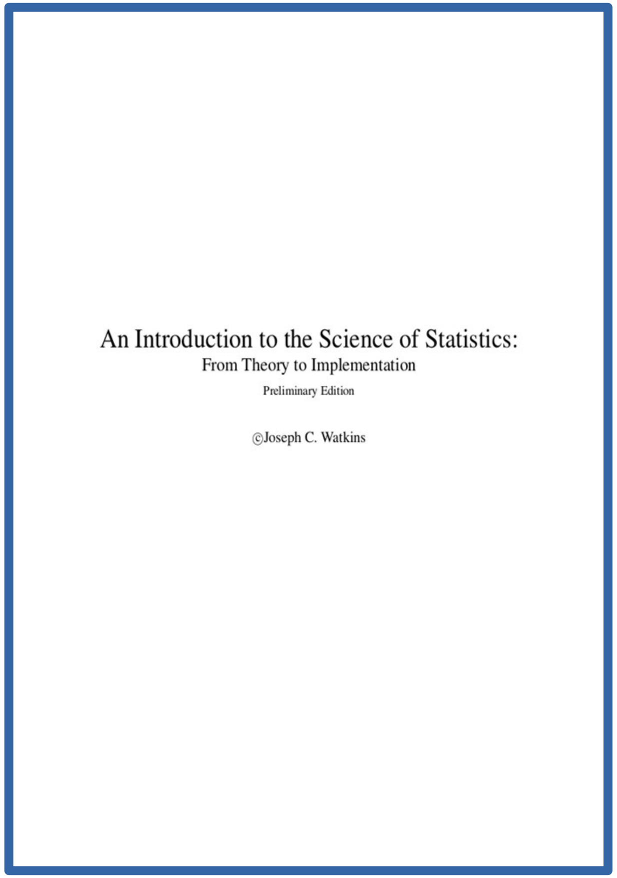
Queen Mary
University of London

- Week 1: Introduction to Probabilities
- Week 2: More Probabilities
- Week 3: Descriptive statistics and Visualisation
- Week 4: Estimators and Limit Laws
- Week 5: Maximum Likelihood Estimators
- Weeks 6 & 8: Hypothesis Testing
- Week 9: Independence, Dependence and Correlation
- Week 10: Linear Regression
- Week 11: Special Topic

REFERENCE BOOKS FOR THE MODULE



E-copy from the library



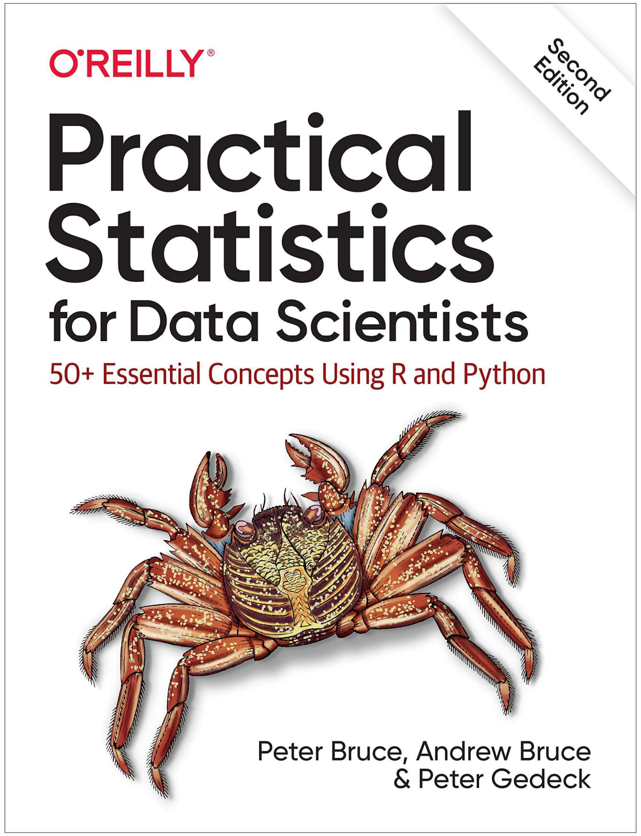
An Introduction to the Science of Statistics:

From Theory to Implementation

Preliminary Edition

©Joseph C. Watkins

https://www.math.arizona.edu/~jwatkins/statbook.pdf



WARNING.

Queen Mary
University of London

- I don’t teach statistics and probabilities in the “standard way”
- You will not hear me talk much about random variables
- I think they create confusion and prevent a good understanding
- If you want to know what they are, check out standard textbooks

PROGRAMMING IN THE MODULE

- We will be using **Python 3**!
  - And a number of packages like numpy and matplotlib.

PROGRAMMING IN THE MODULE

Queen Mary University of London

- Python code accompanies the lectures
- Coding breaks will be represented by python™ at the bottom of slides
- Code can be found at

  https://hub.comp-teach.qmul.ac.uk
- Sign in and choose Applied Statistics
- This will give you access to Jupyter notebooks for the lectures

DATA ANALYSIS: METHODOLOGY

Queen Mary University of London

1) Set your objectives
   - Modelling random process
   - Forecasting
   - Testing statistical hypothesis
2) Collect data and assess quality according to objectives
3) Get a feeling of the data: *Descriptive Statistics, Visualisation*
4) Select methods: *Distribution Fitting, Hypothesis Testing, Regression Analysis*
5) Carry out analysis: *Python*
6) Compare findings with objectives and/or expectations

SOME PUBLICLY AVAILABLE DATA

- OECD data, e.g. economics, demographics, agriculture, health
  - http://stats.oecd.org/
- United Nations data, e.g. development indicators, crime, health, trade
  - http://data.un.org
- Gov.uk open data, e.g. government, transportation, education, health
  - https://www.data.gov.uk
- Kaggle: big repository of data for ML projects
  - https://www.kaggle.com

CALCULUS

- The calculus in these lectures will be very basic. To follow you just need to be able to:
  - Differentiate elementary functions (e.g. $x^n, e^{-\frac{x^2}{2}}, \ldots$)
  - Integrate elementary function
- If you need a refresher:
  - MIT open access courses (choose Single Variable Calculus): https://ocw.mit.edu/search/?d=Mathematics&s=department_course_numbers.sort_coursenum
  - Important formulas and some exercises: https://www.mathcentre.ac.uk/resources/uploaded/final0502-calc-ref-ukmlsc.pdf

PROBABILITY THEORY

A brief introduction

Why is probability theory important?

Why is probability theory important?

- Natural sciences
  - Ecology, biology, chemistry, biochemistry: reactions modelled as happening randomly
  - Physics: quantum mechanics is founded on randomness.

  As far as we know, the universe is intrinsically random. "God" does play dice.
- Engineering
  - Many machine-learning algorithms use probabilities – e.g. ChatGPT
  - Optimisation problems used random algorithms
- Maths
  - Probabilities used outside of probability theory – e.g. number theory, differential equations
- Finance
  - All derivative products are priced using ideas from probability theory

PRELIMINARIES

PRELIMINARIES: SETS

Queen Mary
University of London

1) Writing down sets:

- Finite sets can be written explicitly by listing their elements:

  $\{2,3,5,7,11\}$ or $\{\texttt{true}, \texttt{false}\}$ or $\{a, b, aa, bb, ab, ba\}$

- The empty set contains no element and is written $\emptyset$
- The order in which a set is written is irrelevant

$$\{2,3,5,7,11\} = \{11,2,3,7,5\}$$

- This distinguishes sets from tuples

$$(2,3,5,7,11) \neq (11,2,3,7,5)$$

PRELIMINARIES: SETS

1) Writing down sets:

- Infinite sets must obviously be written differently
- Special names for special sets:

Symbol	Meaning	(Some) Elements
$\mathbb{N}$	Natural numbers	$\{0,1,2,3,4,\ldots\}$
$\mathbb{Z}$	Integers	$\{\ldots,-2,-1,0,1,2,\ldots\}$
$\mathbb{Q}$	Rational numbers	$\left\{\ldots -1,-\frac{1}{2},-\frac{1}{4},0,\frac{1}{3},2,\ldots\right\}$
$\mathbb{R}$	Real numbers	$\left\{\ldots -\sqrt{2},\pi,\frac{1}{e},1,\frac{1}{3},\ldots\right\}$

PRELIMINARIES: SETS

1) Writing down sets:

- The most frequent way of defining/writing down a set is to restrict a previously defined set using a predicate (logical condition)
- Examples

  $\{x \in \mathbb{N} \mid 5 < x\}$ Read: *All natural numbers* x *such that* $5 < x$

  $\{x \in \mathbb{Z} \mid x \ (mod\ 2) = 0\}$ Read: *All integers* x *such that* x *is divisible by 2*
- This kind of notation will feature prominently in this course, make sure you understand it!

PRELIMINARIES: SETS

2) Operations on sets:

1. **Intersection.** The intersection $A \cap B$ contains precisely the elements which belong both to A *and* to B :

$$\{a, b, aa, bb, ab, ba\} \cap \{a, c, aa, cc, ac, ca\} = \{a, aa\}$$

2. **Union.** The union $A \cup B$ contains precisely the elements which belong *either to* A *or* to B (or both!):

$$\{a, b, aa, bb, ab, ba\} \cup \{a, c, aa, cc, ac, ca\} = \{a, b, c, aa, bb, cc, ab, ba, ac, ca\}$$

3. **Relative complement/set difference.** The relative complement $A \setminus B$ contains precisely the elements of A which do not belong to B:

$$\{a, b, aa, bb, ab, ba\} \setminus \{a, c, aa, cc, ac, ca\} = \{b, bb, ab, ba\}$$

PRELIMINARIES: SETS

2) Operations on sets:

4. **Cartesian product.** The Cartesian product $A \times B$ contains precisely the *pairs* of elements such that the first element belongs to A and the second elements belongs to B

$$\{\texttt{true}, \texttt{false}\} \times \{1,2\} = \{(\texttt{true}, 1), (\texttt{true}, 2), (\texttt{false}, 1), (\texttt{false}, 2)\}$$

Essential to understand repeated experiments!

PRELIMINARIES: SETS

3) Membership and inclusion:

- To say that an element belongs to a set we use the membership symbol

$$\texttt{true} \in \{\texttt{true}, \texttt{false}\}$$

- Its negation is written

$$7 \notin \{\texttt{true}, \texttt{false}\}$$

- A is included in B, notation $A \subseteq B$, if every element of A is an element of B. In other words, if $x \in A \Rightarrow x \in B$

PRELIMINARIES: FUNCTIONS

Queen Mary
University of London

4) Functions:

- A function $f: X \to Y$ is a relation between X and Y, i.e. a subset of $X \times Y$
- This relation must satisfy the following condition:
  1. If $(x, y) \in f$ and $(x, y') \in f$ then $y = y'$
  2. For every $x \in X$ there exists a $y \in Y$ such that $(x, y) \in f$
- In other words, each "input" $x \in X$ is related to exactly *one* "output" $y \in Y$
- We therefore write this y as $f(x)$
- Functions are written down in the following format

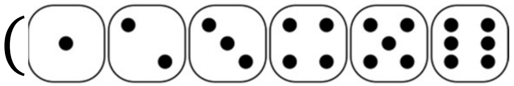
$$f: X \to Y, x \mapsto f(x)$$

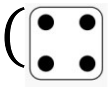
- Example: $f: \mathbb{R} \to \mathbb{R}, x \mapsto x^2$
- Exercise: write down the square root function.

MEASURES

MEASURES: DISCRETE CASE

- Central concept: **Probability Measure**, a special class of **measures**
- In the discrete case, measures generalize **counting**

$count($  $) = 6 = \sum_{i=1}^{6} 1$

$count($  $) = 1 = \sum_{i=1}^{1} 1$

$$count(\{0,1,2,3,4,\dots\}) = \infty = \sum_{i=1}^{\infty} 1$$

MEASURES: TOWARDS A DEFINITION

- Abstract properties of counting
  1. $count(\quad)$ takes a **set** as input and returns a **number**
     for example: $count(\{1,2,3\}) = 3$
  2. $count(\emptyset) = 0$
  3. $count(\{a, b, c\} \cup \{d, e, f\}) = count(\{a, b, c\}) + count(\{d, e, f\})$

  3. *More generally:* $count(A \cup B) = count(A) + count(B)$ *if* $A \cap B = \emptyset$ *(otherwise double-counting!)*

- This is almost the definition of a measure.

MEASURES: CONTINUOUS CASE

- In the continuous case, measures generalise lengths, areas, volumes, etc

$$length([a,b]) = b - a = \int_a^b 1\ dx$$

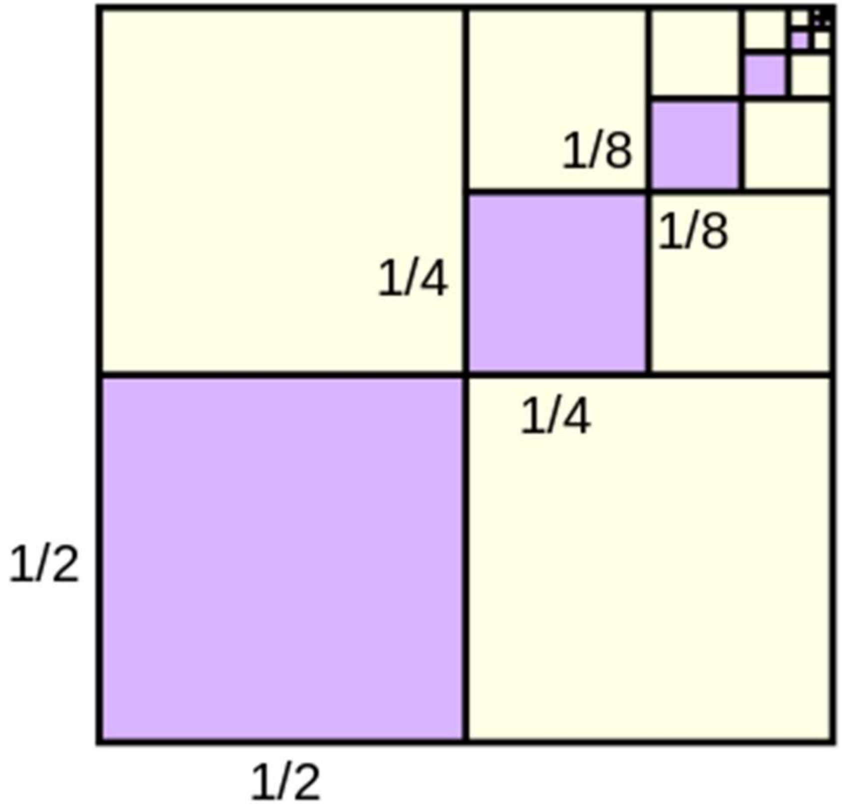
$$area([a,b] \times [c,d]) = (b-a)(d-c) = \int_a^b \int_c^d 1\ dx\ dy$$

$$volume([a,b] \times [c,d] \times [e,f]) = \int_a^b \int_c^d \int_e^f 1\ dx\ dy\ dz$$

$$length\big((-\infty, \infty)\big) = \infty$$

MEASURES: CONTINUOUS CASE

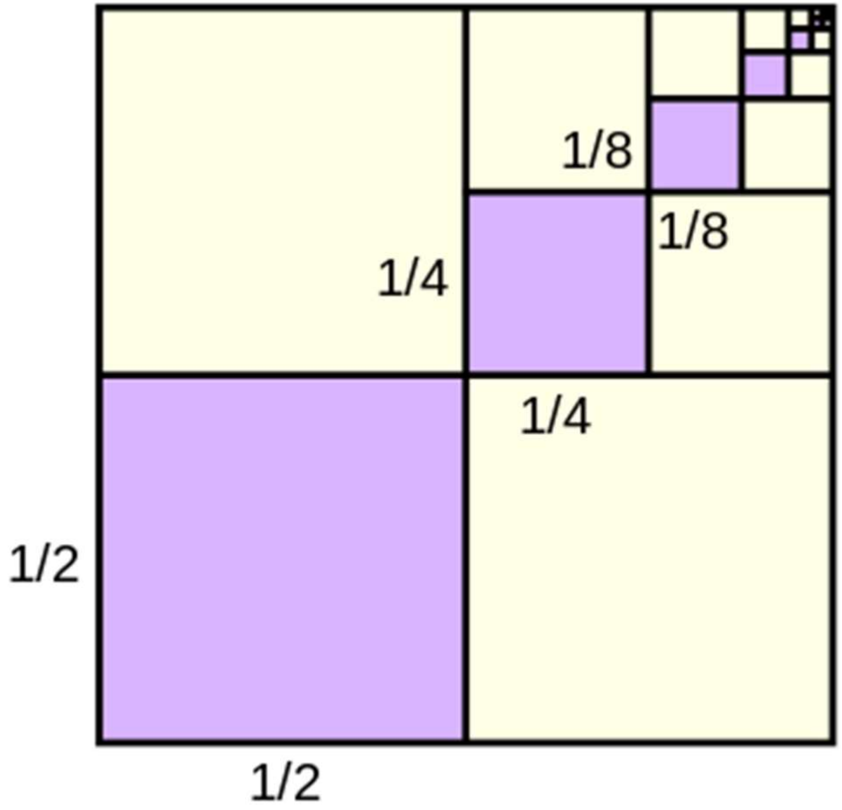
- What is the area coloured in purple?



MEASURES: CONTINUOUS CASE

- What is the area coloured in purple?
- The area coloured in purple is given by

$$\frac{1}{4}+\frac{1}{16}+\frac{1}{64}+\cdots=\frac{1}{3}$$



MEASURES: DEFINITION

- Abstract properties of area
  1. $area(\quad)$ takes a **set** as input and returns a **number**

     for example: $area\left(\left[0,\frac{1}{2}\right]\times\left[0,\frac{1}{2}\right]\right)=\frac{1}{4}$
  2. $area(\emptyset)=0$
  3. $area(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}area(A_i)$ *if all* A_is *are pairwise disjoint (otherwise double-counting!). See example on previous slide.*

- This is also true for length and volumes
- In fact, this the definition of a measure

MEASURES: DEFINITION

- A measure $\mathbb{M}$ on a set A is a function taking subsets $V \subseteq A$ as inputs and returns a real number $\mathbb{M}(V) \in \mathbb{R}$
- We will often think of $\mathbb{M}(V)$ as the "mass" of V
- $\mathbb{M}$ must satisfy the conditions on the previous slides
  1. $\mathbb{M}(V) \geq 0$
  2. $\mathbb{M}(\emptyset) = 0$
  3. $\mathbb{M}(\bigcup_{i=1}^{\infty} V_i) = \sum_{i=1}^{\infty} \mathbb{M}(V_i)$ (σ-additivity)

Note: in good books you will see that measures are not defined on *all* subsets but only on a certain type of collection of subsets called a σ-algebra. I will not discuss this subtlety; I only want you to remember that measures take subsets as inputs.

PROBABILITY MEASURES

- **Probability** measures are measures which are **normalized:**
  The measure of the entire set must be 1, formally $\mathbb{M}(A) = 1$.
- Examples:

$\mathbb{P}($⚀⚁⚂⚃⚄⚅$) = 1 = \sum_{i=1}^{6} \frac{1}{6}$

$\mathbb{P}($⚃$) = \frac{1}{6} = \sum_{i=1}^{1} \frac{1}{6}$

$$\mathbb{P}([a, b]) = 1 = \int_a^b \frac{1}{(b-a)}\, dx$$

DENSITIES: discrete case

- Examples so far: every “point” contributes equally to the total measure
- What if some points contribute more/are “heavier” ?

density

$$Mass(\text{2kg}, \text{1kg}, \text{3kg}) = 2 + 1 + 3 = \sum_{i=1}^{3} m_i . 1$$

We’re still counting!

DENSITIES: continuous case

Queen Mary
University of London

- Imagine a rod of length 4 m.
  - Along 1st meter it weights 1kg/m
  - Along 2nd meter it weights 2kg/m
  - Along 3rd meter it weights 3kg/m
  - Along 4th meter it weights 4kg/m
- What is the weight of the entire rod?

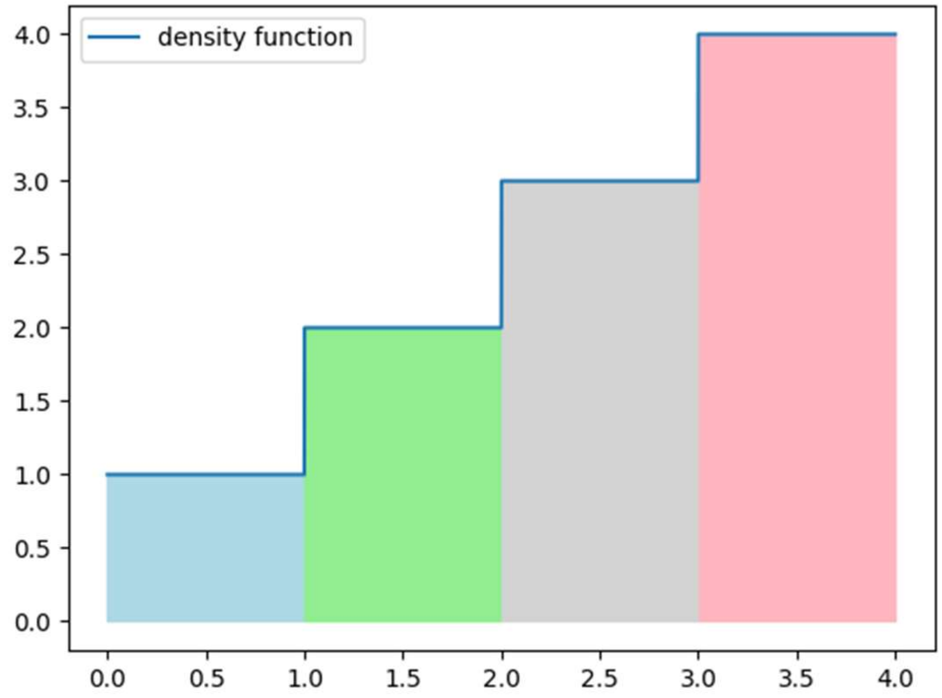
DENSITIES: continuous case

- Imagine a rod of length 4 m.
  - Along 1st meter it weights 1kg/m
  - Along 2nd meter it weights 2kg/m
  - Along 3rd meter it weights 3kg/m
  - Along 4th meter it weights 4kg/m
- What is the weight of the entire rod?

1kg+2kg+3kg+4kg=10kg

$$\text{Weight} = \int_0^4 density(x).1\, dx$$

Standard length is reweighted



DENSITIES: continuous case

- Imagine now a rod which gets heavier as you move away from one of the ends. Formally, imagine length l and density x kg/m at x meter from the end

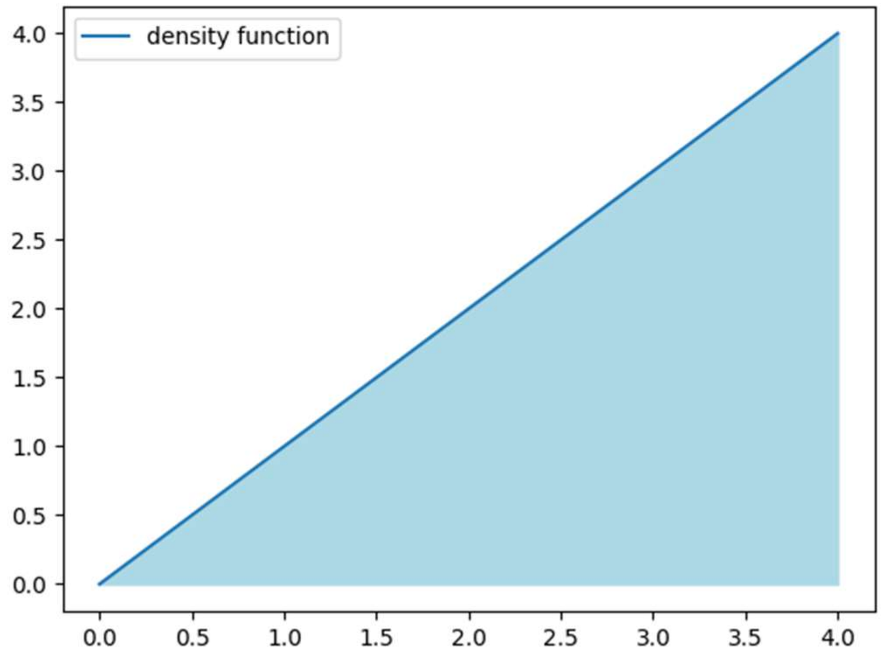
$$Mass(Rod) = \int_0^l x.1\,dx = \frac{l^2}{2}$$

density

Still a kind of length!

$$Mass(First\ Half\ Rod) = \int_0^{l/2} x\,.1\,dx$$

$$= \frac{l^2}{8}$$



PMF and PDF

- We can do the same thing with **probability** measures – it just needs to add up/integrate to 1
- Discrete case: loaded dice.

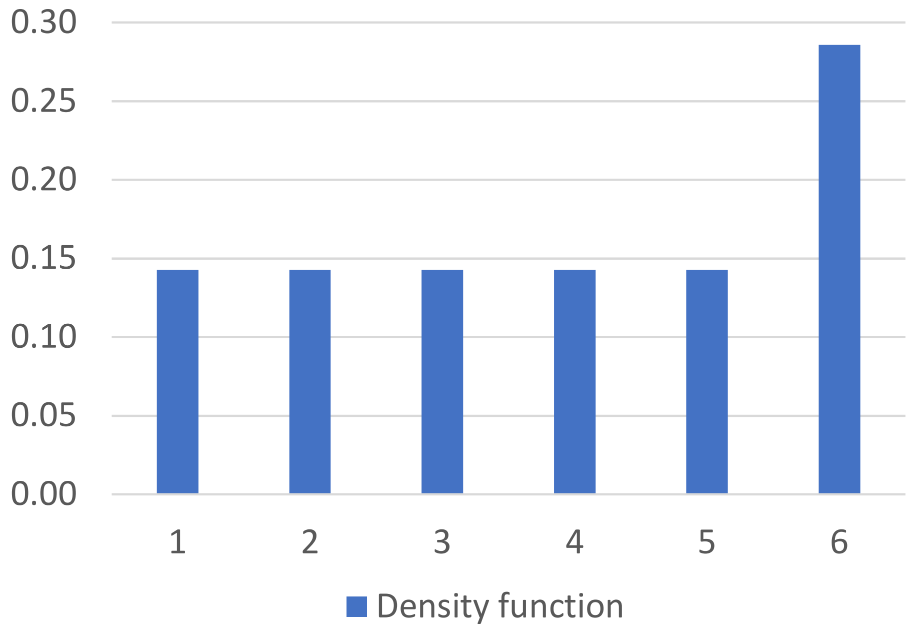
$\mathbb{P}($  $) = \frac{2}{7}$

$\mathbb{P}(other\ faces) = \frac{1}{7}$

- Probability Mass Function (PMF)

$$f : \{1,2,3,4,5,6\} \to \mathbb{R}^+$$

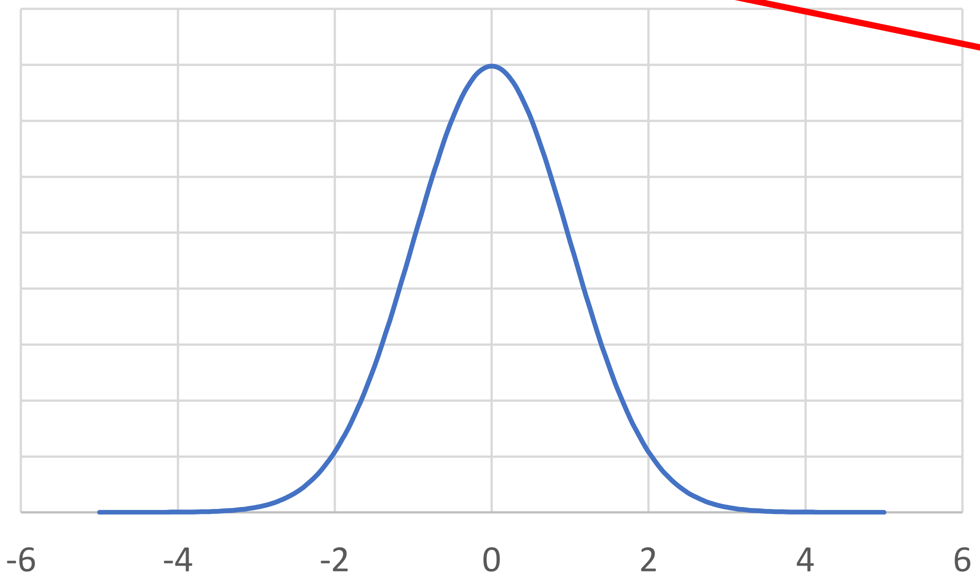
$$1,2,3,4,5,6 \mapsto \frac{1}{7} \qquad 6 \mapsto \frac{2}{7}$$



PMF and PDF

- Continuous case: normal distribution

$$\mathbb{P}\big((-\infty, t]\big) = \int_{-\infty}^{t} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}_{} .1\, dx$$



Probability Density Function (PDF). Integrates to 1.

PMF and PDF

Queen Mary
University of London

- In the discrete case:

Probability Density/Mass Function tells us by how much counting (measure) is reweighted

- In the continuous case:

Probability Density Function tells us by how much the usual integration/length (dx) is reweighted

- Theoretically, probability theory is about **probability measures**. In practice, Probability Density Functions – **PDF**s - are the most useful mathematical objects.

PMF and PDF: USAGE

- Practically, the density gives us a concrete way to compute the probability mass of subsets
- Discrete case: $\mathbb{P}(A) = \sum_{x \in A} f(x)$

Eg: $Binom(10, 0.4)(\{1,2,3\}) = f(1) + f(2) + f(3)$

$$= \binom{10}{1} 0.4^1 0.6^9 + \binom{10}{2} 0.4^2 0.6^8 + \binom{10}{3} 0.4^3 0.6^7$$

- Continuous case: $\mathbb{P}(A) = \int_{x \in A} f(x) dx$

Eg: $N(0,1)([2,3]) = \int_2^3 f(x) dx = \int_2^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

CUMULATIVE DISTRIBUTION FUNCTION

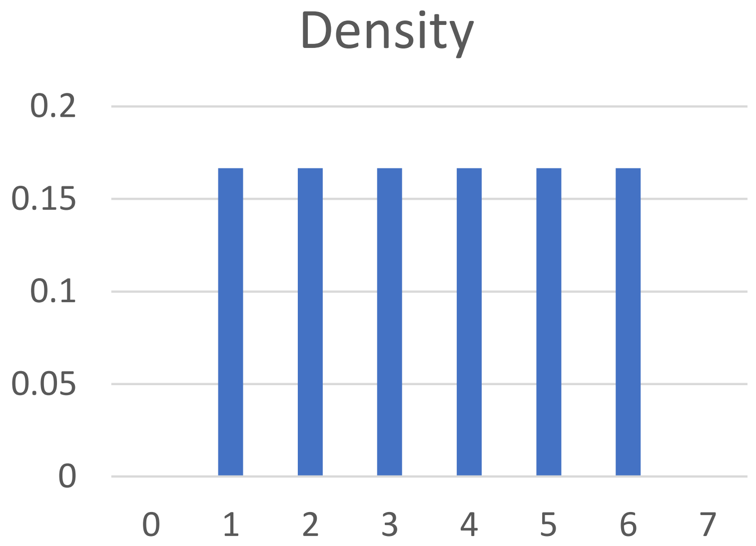
- For distributions on $\mathbb{R}$ it always makes sense to define

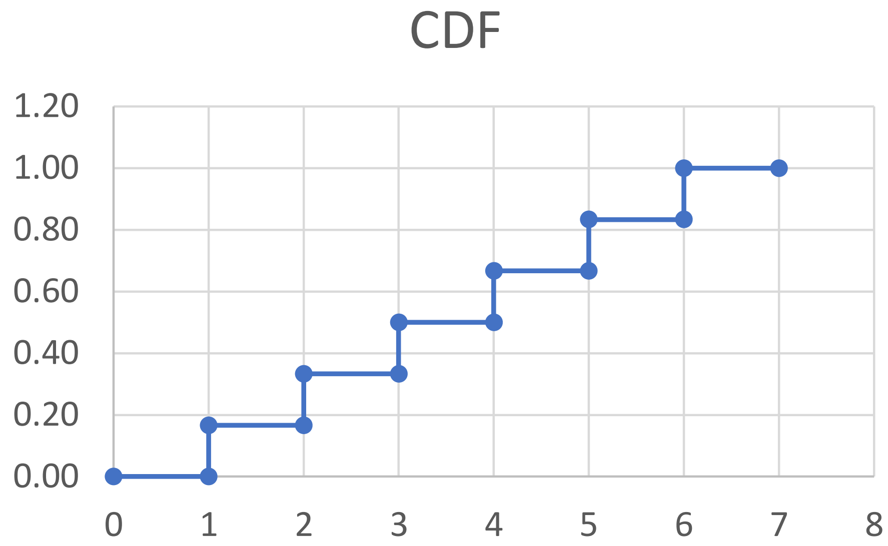
$$F(t) = \mathbb{P}\big((-\infty, t]\big), t \in \mathbb{R}$$

- This function is called the **Cumulative Distribution Function** or **CDF**

- In terms of the density function we have
  - Discrete case: $F(t) = \sum_{x \le t} f(x)$
  - Continuous case: $F(t) = \int_{-\infty}^{t} f(x)dx$

CDF: FAIR DICE





PDF vs CDF

- ***Fundamental Theorem of Calculus:***

If $f: \mathbb{R} \to \mathbb{R}$ is continuous and integrable and $F(t) = \int_{-\infty}^{t} f(x)dx$ then

$$\frac{\partial}{\partial t} F(t) = f(t)$$

- If f is a PDF, then F is a CDF
- The fundamental theorem of calculus allows us to compute a density from a CDF, i.e. from $\mathbb{P}\big((-\infty, t]\big)$
- This is extremely useful in practice