Introduction to Modelling and Analysis of Data (MAD)

Bulat Ibragimov
IMAGE section,
Department of Computer Science (DIKU)

UNIVERSITY OF COPENHAGEN

About Us

- Lecturers:

Bulat Ibragimov
(course responsible)
bulat@di.ku.dk

Kim Steenstrup Pedersen
kimstp@di.ku.dk

- Teaching Assistants:
  - MAD Hold 1: Tudor-Laurentiu Dascalu
  - MAD Hold 2: Rong Fu
  - MAD Hold 3: Christian Feveile Holck Lohman
  - MAD Hold 4: Jack Henry Monaghan
  - MAD Hold 5: Zeb Harris Buus Nielsen
  - MAD Hold 6: Nichlas Langhoff Rasmussen
  - MAD Hold 7: Christian Charlie Virt

Tentative Schedule

- Week 1: Linear regression
- Week 2: Non-linear regression and statistics
- Week 3: Advanced statistics and Maximum Likelihood approach to parameter estimation and regression
- Week 4: Bayesian perspective on regression and Principal Component Analysis (PCA)
- Week 5: Classification
- Week 6: Sampling
- Week 7: Clustering

Recommended Academic Qualifications

- From course description:
  "Mathematical knowledge equivalent to those obtained in the courses LinAlgDat, DMA, and MASD or similar. Basic knowledge of programming."
- MAD is a partner course to MASD (Mathematical Analysis and Statistics for Computer Scientists), block 1.
  1. MASD focused on mathematical analysis and probability theory that is fundamental to data science (basics).
  2. MAD will turn towards more advanced statistics and machine learning
- MAD builds on the probability theory and calculus from MASD.
- MAD also relies heavily on linear algebra.
- We assume you know how to write programs (in some programming language). It's a tool that we will be using in MAD.

Your Background

- You come from several educations – computer science, Physics, Bioinformatics, …
- And your background knowledge varies

- Need a probability theory brush-up? We have provided some suggested reading in Absalon: Blitzstein and Hwang Ch.: 1, 2, 3.1 - 3.3, 3.5 - 3.8, 4.1 - 4.6, 5.1 - 5.4, 7.5.
- New to Python? Start doing the optional exercise on Python.

Your Background: Survey results

Which education are you enrolled in?

Computer Science	96 respondents	**91 %** ✓
Physics	8 respondents	8 %
Mathematics	2 respondents	2 %
Bioinformatics		0 %
Biology		0 %
None of the above		0 %

Attempts: 106 out of 106

How far are you in your studies?

I'm taking this course as part of my bachelors studies	105 respondents	**99 %** ✓
I'm taking this course as part of my MSc studies	1 respondent	1 %
I'm taking this course as part of my PhD		0 %
None of the above		0 %

Attempts: 105 out of 106

Have you followed the MASD course?

Yes	87 respondents	**82 %** ✓
No	18 respondents	17 %
No Answer	1 respondent	1 %

Your Background: Survey results

Have you followed a course on probability theory and statistics?

Yes	16 respondents	**15 %**
Yes, I followed MASD	85 respondents	80 %
No	5 respondents	5 %

Attempts: 106 out of 106

How would you describe your experience with the Python programming language?

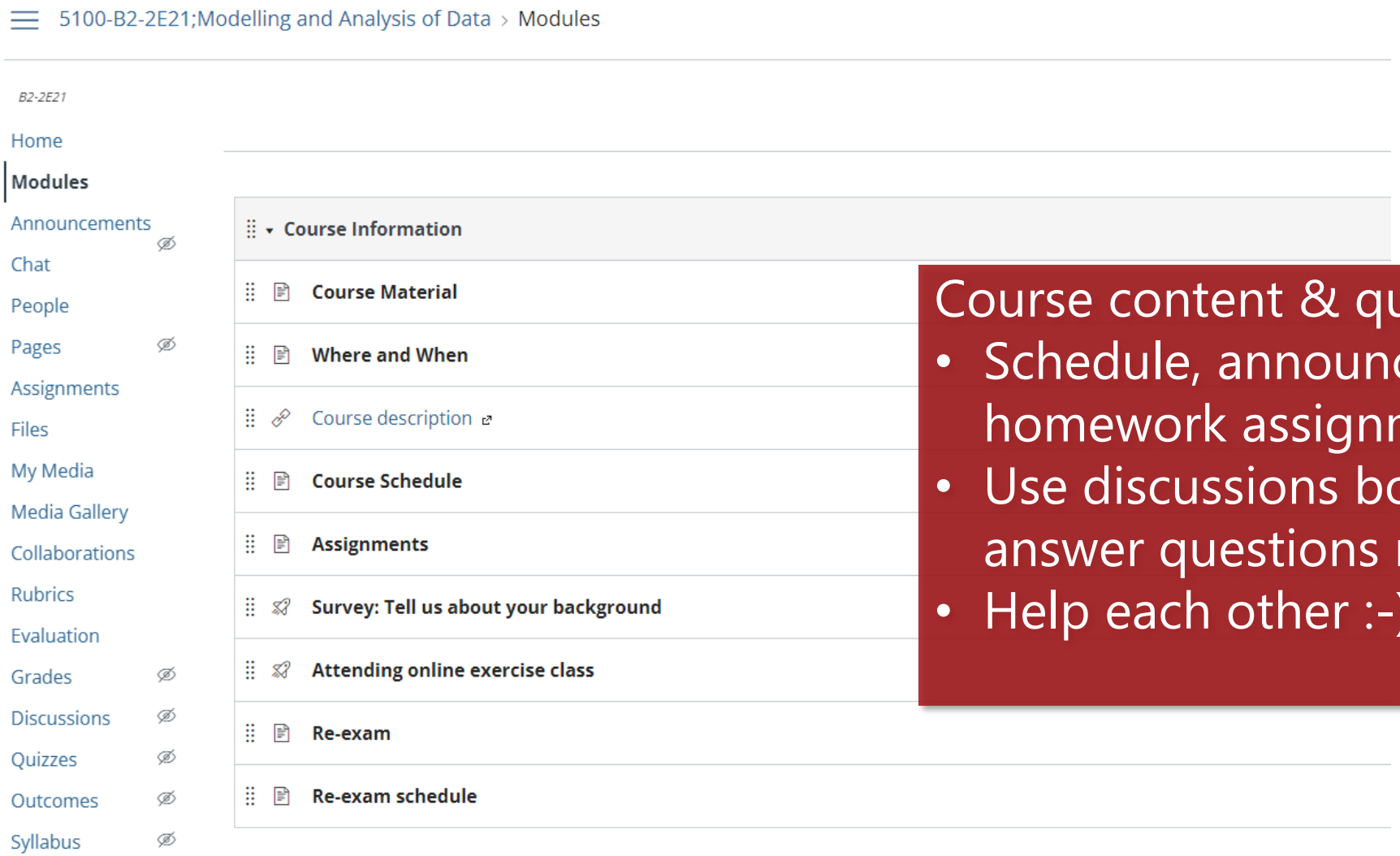
Don't know the language	12 respondents	**11 %**
Basic knowledge - I have written a couple of lines of code	53 respondents	50 %
Medium level - I have written scripts / modules or Jupyter notebooks from scratch	28 respondents	26 %
High level - I have written complex programs consisting of several modules	11 respondents	10 %
I'm an expert	2 respondents	2 %

Attempts: 106 out of 106

Have you worked with Jupyter notebooks before?

Yes	79 respondents	**75 %**
No	27 respondents	25 %

Course Organization: Absalon



Course content & questions:

- Schedule, announcements, lectures, homework assignments, . . .
- Use discussions board to ask and answer questions related to MAD!
- Help each other :-)

Course Organization: When and Where

- Lectures:
  - Tuesdays 10:00-11:45 – aud-AUD 01 AKB, Universitetsparken 13.
  - Thursdays 10:00-11:45 - aud-Aud 01 HCØ, Universitetsparken 5, HCØ
- Exercise classes:
  - Thursdays 13:00 – 15:00. Please check the location for your group
- Videos
  - Some videos will be uploaded but not for all the lectures, and they will be not as eleborated as the physical lectures

Assignments (tentative)

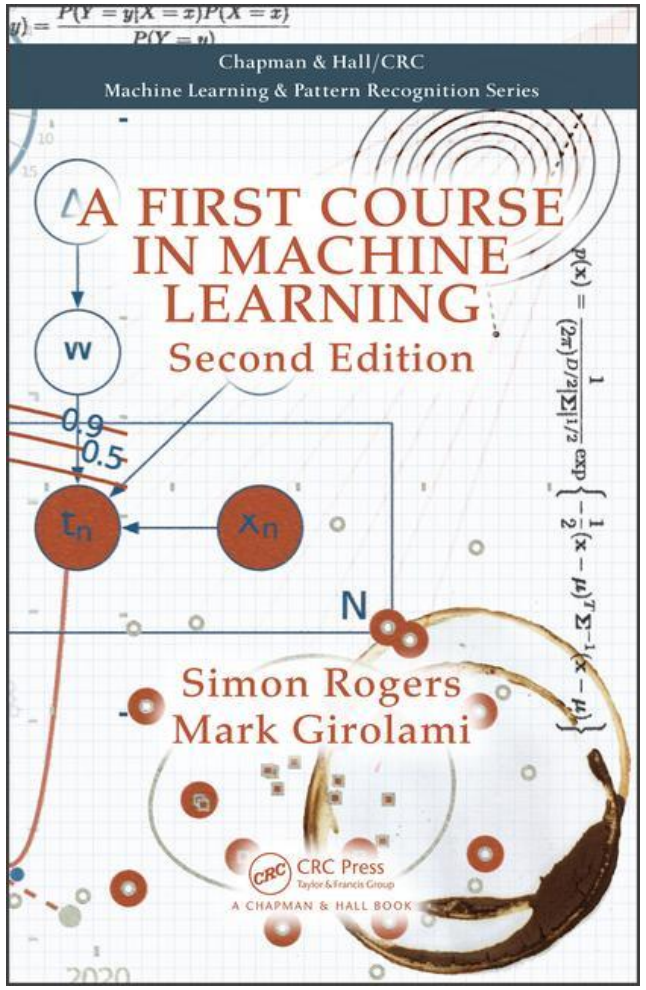
- There will be six assignments. The assignments will be handed out Monday morning (around 10:00) and will have to be handed in **1 week later by Monday night, 23:59**.
  - A1 (22.11.2021 - 29.11.2021)
  - A2 (29.11.2021 - 06.12.2021)
  - A3 (06.12.2021 - 13.12.2021)
  - A4 (13.12.2021 - 20.12.2021)
  - A5 (03.01.2022 - 10.01.2022)
  - A6 (10.01.2022 - 17.01-2022)
- See each assignment for details and potential changes.

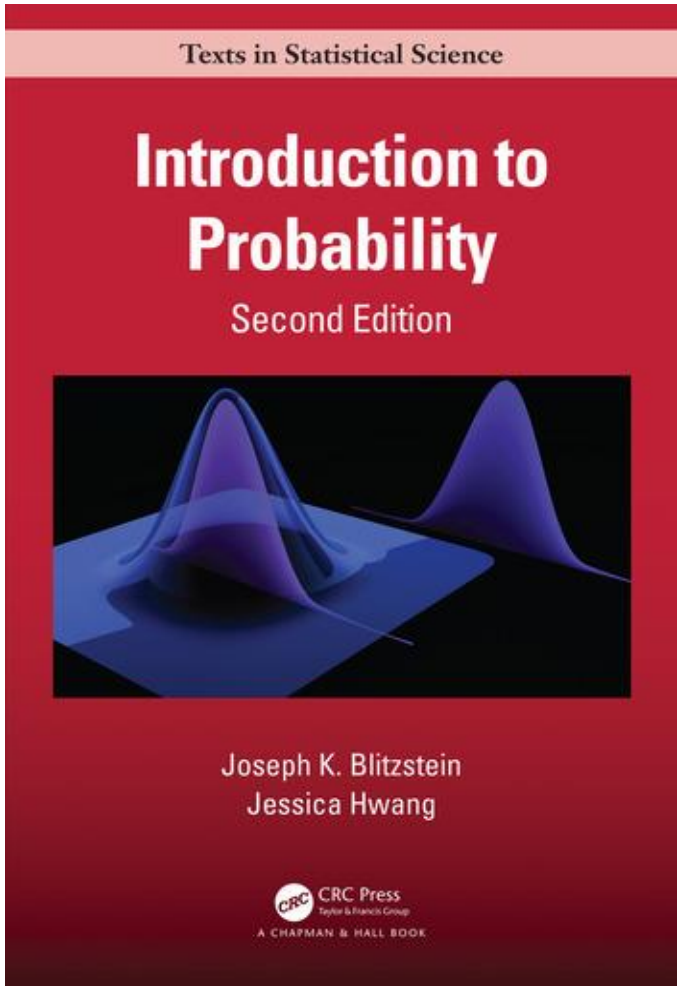
Exam Qualification & Exam

- To qualify for the exam:
  - **All but one of the assignments must be passed in order to be eligible for the exam.** In general, passing means to get 40% of the points per assignment.
  - The assignments must be completed and written individually. But you are allowed (and encouraged) to discuss the exercises in small groups.
  - Don't copy code or text from each other. This will be considered plagiarism.

- Exam:
  - The exam is a 7 days take-home exam (calendar week 3, tentative dates 17.01.2022 – 23.01.2022)
  - **The exam is individual and you are not allowed to work/discuss with each other.**

Course material:





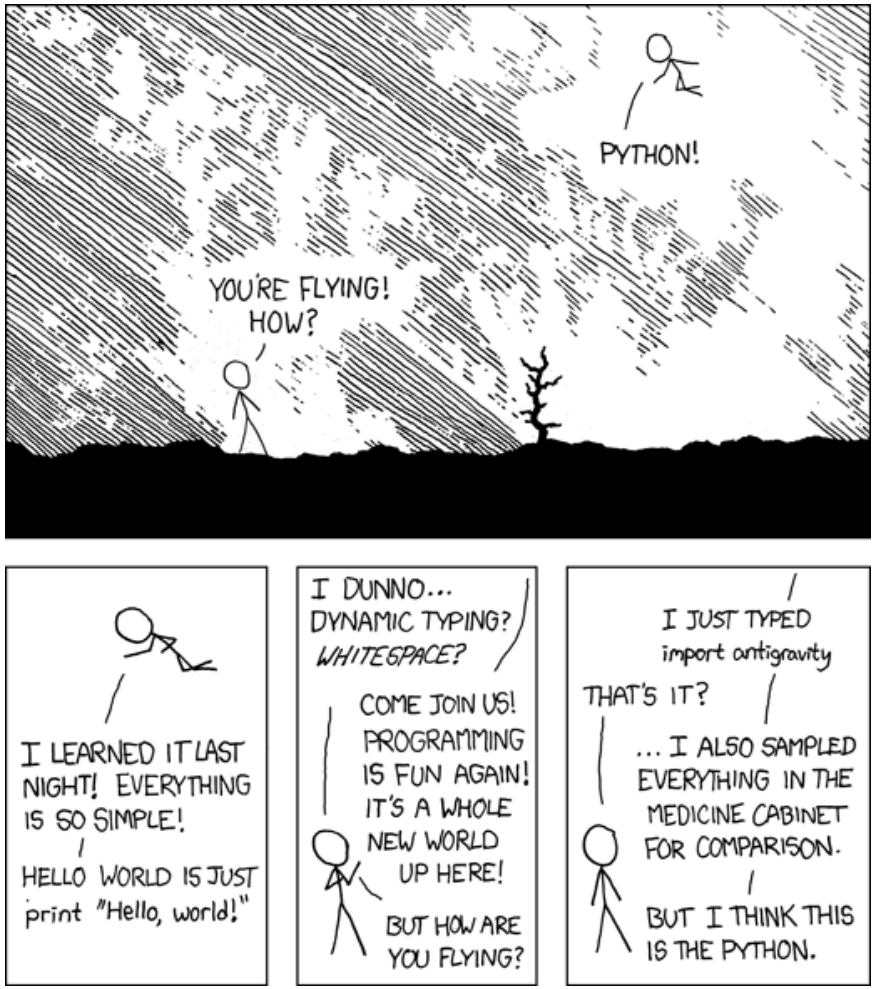
Electronic version at http://probabilitybook.net

Additional material will be available in Absalon

We will be using Python

Its easy to learn!

It's the language of choice in Data Science!

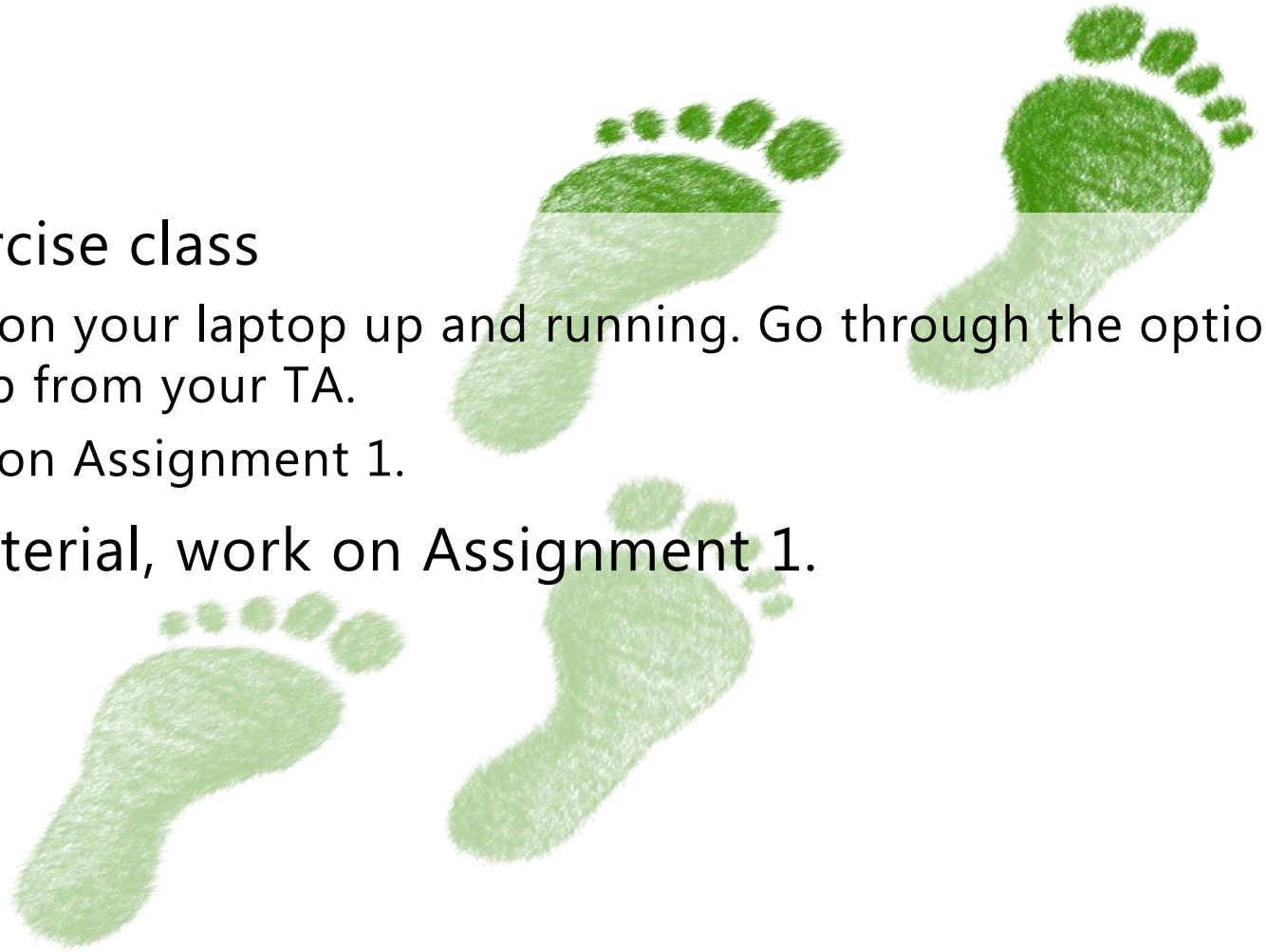


https://xkcd.com/353/

We will be using Python and Jupyter

- We recommend installing Python via the Anaconda distribution.
- If you are new to the language, then solve the optional extra exercise questions.
- We will also be using these packages:
  - Numpy (and Scipy)
  - Matplotlib
  - Scikit-learn (when we allow you to use this)
- For some lectures and assignments, we also use Python Jupyter notebooks. See how to open a Jupyter notebook at https://docs.anaconda.com/anaconda/user-guide/getting-started/ or see https://jupyter.readthedocs.io/en/latest/running.html
- Lets look at an example – notebookexample.ipynb

Next steps



- Join the exercise class
  - Get Python on your laptop up and running. Go through the optional extra exercise and get help from your TA.
  - Get started on Assignment 1.
- Read the material, work on Assignment 1.

Questions?