

Information Theory for Data Science

Assignment 2 – Part A

Prof. Giorgio Taricco
Politecnico di Torino – DET

Assignment 2 rules

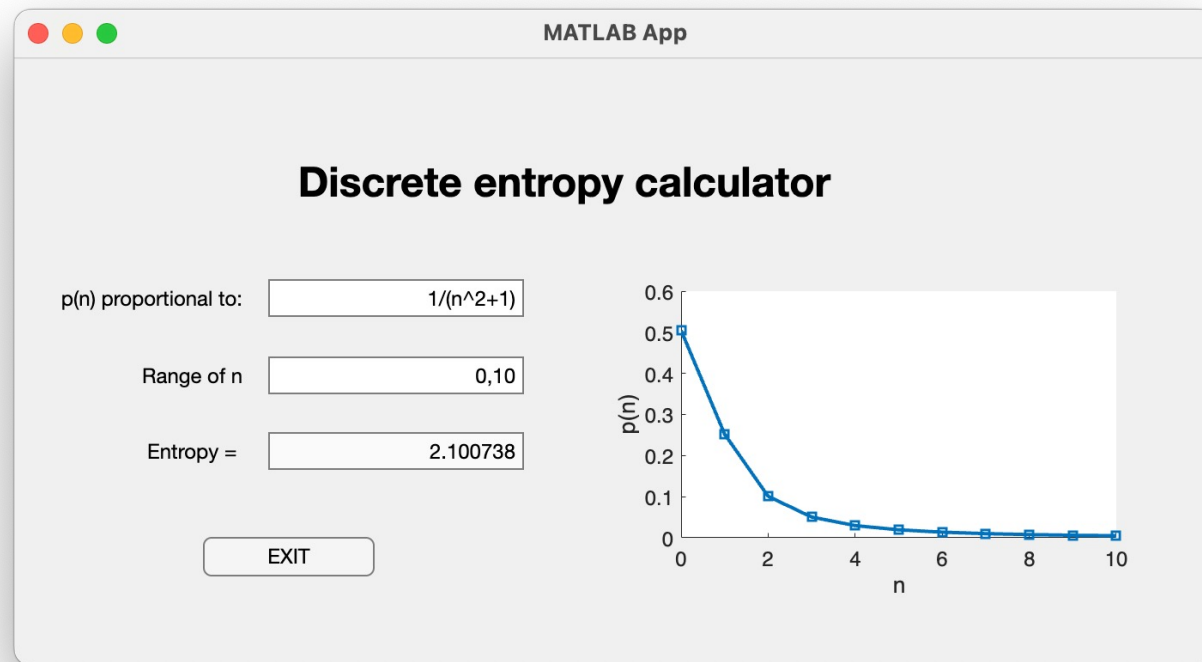
- No cooperation outside of the group is allowed
- Reports must describe the code, the numerical results, and all the analytic steps when required
- Reports must be well written and organized, provide all the implementation details
- Quality is an important element for the evaluation
- Matlab language is preferred (with AppDesigner for GUIs)
- Notebook Python is a (not recommended) backup option which must be well documented for the code and the running instructions and must be runnable as a GUI equivalent to an AppDesigner GUI

Delivery rules (Assignment 2)

- All the files relevant to the assignment (part a and b together) must be collected in a single compressed ZIP file
- The file must be uploaded through POLITO's WEB PORTAL
- Emails are not considered
- The ZIP files must have a name "group<number>.zip"
- The group number will be published before the time you can submit the assignment
- The report must be prepared in latex and saved in PDF format with name "group<number>.pdf"

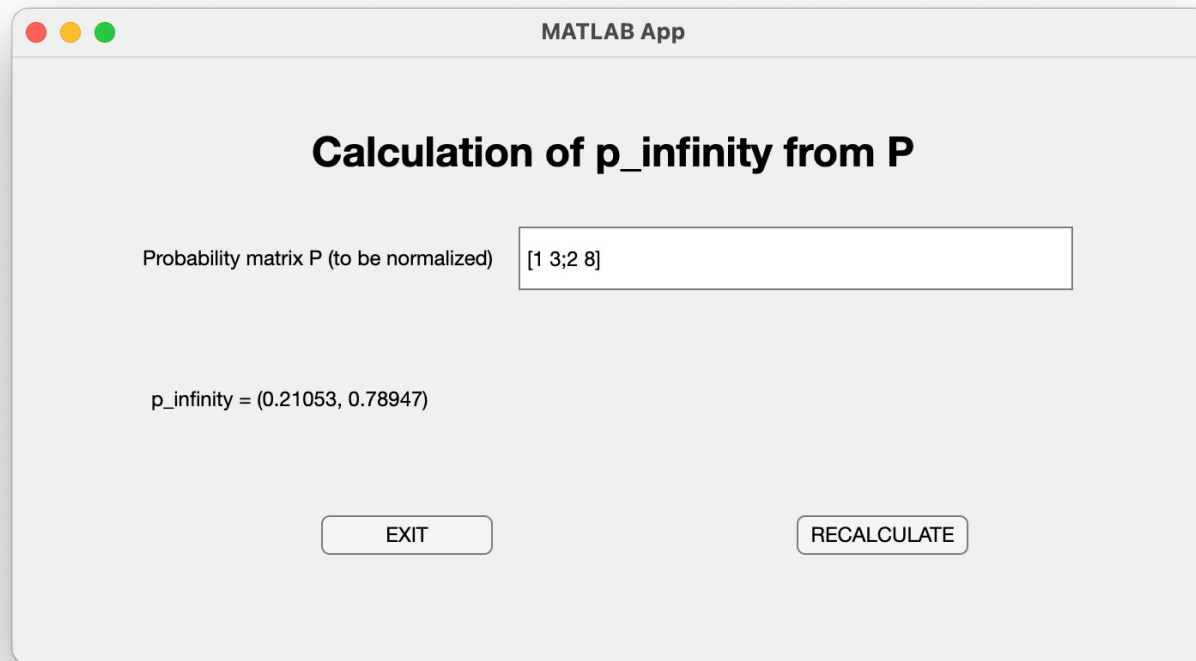
Problem 1

- Make a GUI for the evaluation of the entropy of a discrete RV



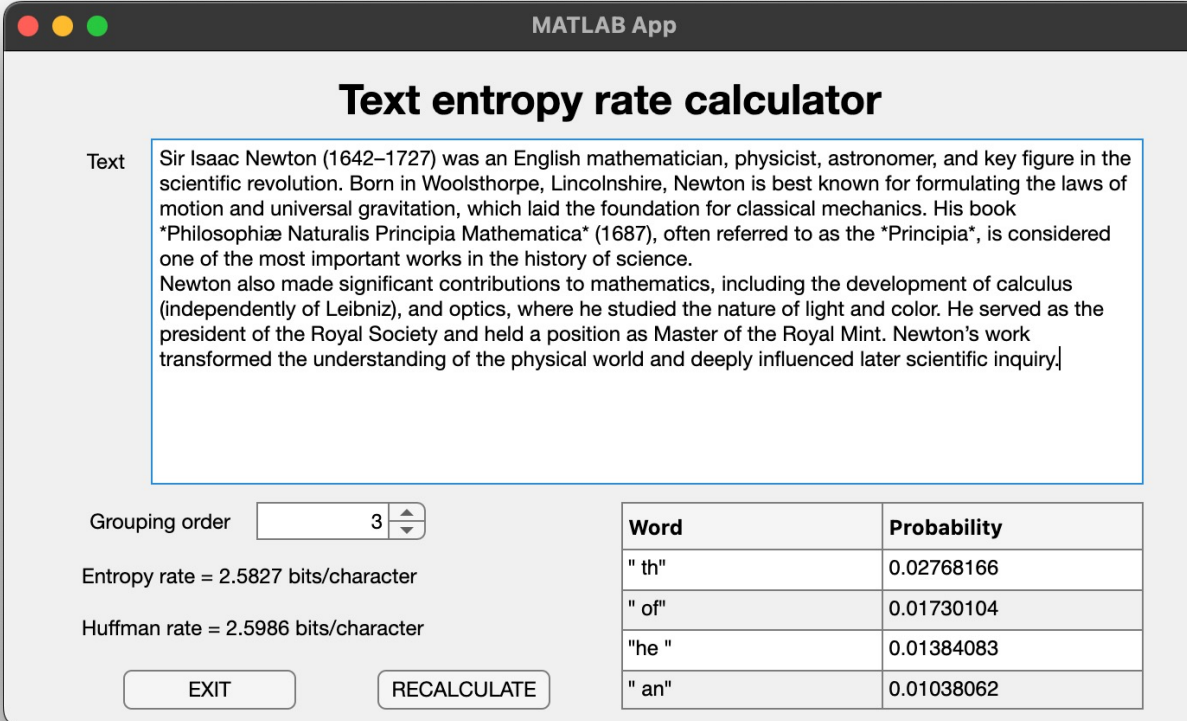
Problem 2

- Make a GUI to calculate \mathbf{p}_{∞} for a transition probability matrix \mathbf{P}



Problem 3

- Make a GUI to enter an ASCII text in an edit box and calculate the entropy rate and the average number of bits per symbol (based on symbol frequencies) by merging consecutive characters in groups of a given order.
- Also output the decreasingly ordered probability distribution



Text entropy rate calculator

Text: Sir Isaac Newton (1642–1727) was an English mathematician, physicist, astronomer, and key figure in the scientific revolution. Born in Woolsthorpe, Lincolnshire, Newton is best known for formulating the laws of motion and universal gravitation, which laid the foundation for classical mechanics. His book *"Philosophiæ Naturalis Principia Mathematica"* (1687), often referred to as the *"Principia"*, is considered one of the most important works in the history of science. Newton also made significant contributions to mathematics, including the development of calculus (independently of Leibniz), and optics, where he studied the nature of light and color. He served as the president of the Royal Society and held a position as Master of the Royal Mint. Newton's work transformed the understanding of the physical world and deeply influenced later scientific inquiry.

Grouping order:

Entropy rate = 2.5827 bits/character

Huffman rate = 2.5986 bits/character

| Word | Probability |
|-------|-------------|
| " th" | 0.02768166 |
| " of" | 0.01730104 |
| "he " | 0.01384083 |
| " an" | 0.01038062 |

Problem 4

- For each exercise provide the analytic solution (if it is possible to find an analytic solution, and include the parameter ranges)
- Otherwise, state that the analytic solution is impossible and provide the numerical solution (the correct choice is part of the evaluation)
- Calculate the entropy of the following distributions:
 1. $p(n) \propto \exp(-\lambda n), n = 0, 1, 2, \dots$
 2. $p(n) \propto \exp(-n^2), n = 0, 1, 2, \dots$
 3. $p(n) \propto n^{-4}, n = 1, 2, 3, \dots$
 4. $p(n) \propto \alpha^n, n = 1, \dots, N$
 5. $p(n) \propto (1 + n^2)^{-k}, n = 0, 1, 2, \dots$ for $k = 1, 2$