Information theory and coding

Guidelines to prepare the written exam of June 2019

The written exam covers the material covered in the course and in the practicals. It is organised in 2 parts: Part 1 concerns the practicals and Part 2 the theory.

Course grading:

- *Practical projects: 30%*
- *Part 1 of written exam: 30%*
- *Part 2 of written exam: 40%*

1 Part 1 (2h) open books

Questions will be in the form of exercises and multiple choice questionnaires.

- Probabilistic modelling of a problem and application information theory to its analysis. Manipulation of information theory quantities (entropies, mutual informations, simple and conditional).
- Construction of trees and poly-trees (Chow-Liu algorithm, propagation of causality basins) based on an oracle about the target joint probability mass function.
- Detection of I-map, D-map, P-map in different graphical probabilistic models.
- Main properties of data compression codes (regular, reversible, prefix-less, complete, instantaneous). Huffman coding, complete and incomplete tree codes. Compression rate.
- Codes for detecting and/or correcting errors on the binary symmetric channel. Repetition codes, linear block codes (Hamming (7, 4)). Information capacity of a (stationary memoryless) channel.

2 Part 2 (2h30) closed books

Below is a list of topics to guide you in preparing this part of the exam, in the form of indicative questions that could be asked.

Algebra of entropies and mutual informations

1. Define mathematically the notions of entropy, conditional entropy, joint entropy, and mutual information.
2. State and graphically represent the algebraic relations among these quantities and their main properties (bounds, inequalities and equalities among quantities), and explain the main steps of the mathematical proofs of these relations.

3. State the chaining rule for joint entropies. Define the notion of conditional mutual information and state the chaining rule for mutual informations. Explain the main steps of the mathematical proofs of these two chaining rules.

4. Define the notion of Markov chain (over three discrete random variables). State, prove, discuss and illustrate the data processing inequality. Give an example where the data processing inequality can not be applied and where it is also not satisfied. State and discuss the corollaries of the data processing inequality.

Source coding theorem and data compression algorithms

1. Define the notion of convergence in probability, formulate the AEP, and explain the principle of its proof. Define the set of ϵ-typical messages of length n ($A_\epsilon^{(n)}$), and state exactly the 4 main properties of this set. Explain the use of this notion in the context of data compression.

2. Define the notion of (discrete) stationary data source. State the 2 definitions of the entropy rate of a discrete stationary source ($H(S)$ and $H'(S)$). Show mathematically why the existence of $H'(S)$ implies that of $H(S)$. Show mathematically that $H'(S)$ exists for a stationary source.

3. State the first Shannon theorem and prove it mathematically. Give two examples of reversible data compression methods and explain their advantages and drawbacks.

Channel coding theorem and error correcting codes

1. Define mathematically the general notion of discrete channel; then explain in this discrete setting the further mathematical restrictions corresponding to the notion of i) causal channel; ii) causal memoryless channel, iii) causal memoryless and stationary channel.

2. State the mathematical definition of the per-symbol information capacity of the discrete, causal and memoryless stationary channel, and explain why and how this quantity depends on the channel properties. Present the binary symmetric channel, and mathematically derive the expression of its information capacity as a function of the error rate p; discuss the resulting expression.

3. Define the notion of (M, n)-code for a channel $(\mathcal{X}, P(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$. Explain decoding rules: i) maximum posterior probability (Bayes rule) and its practical caveat; ii) maximum likelihood and why it is not really restrictive. Define the notion of communication rate R of an (M, n)-code, the notion of achievable rate and the notion of operational capacity.

4. State the second Shannon theorem, and sketch its proof in the context of a discrete causal memoryless and stationary channel. Discuss the generalisation to continuous channels and to channels with correlated noise.

5. Explain the main stakes of channel coding and decoding, in terms of algorithmic, geometric and statistical considerations.