Syllabus

Mário S. Alvim
(msalvim@dcc.ufmg.br)

Information Theory

DCC-UFMG
(2017/02)

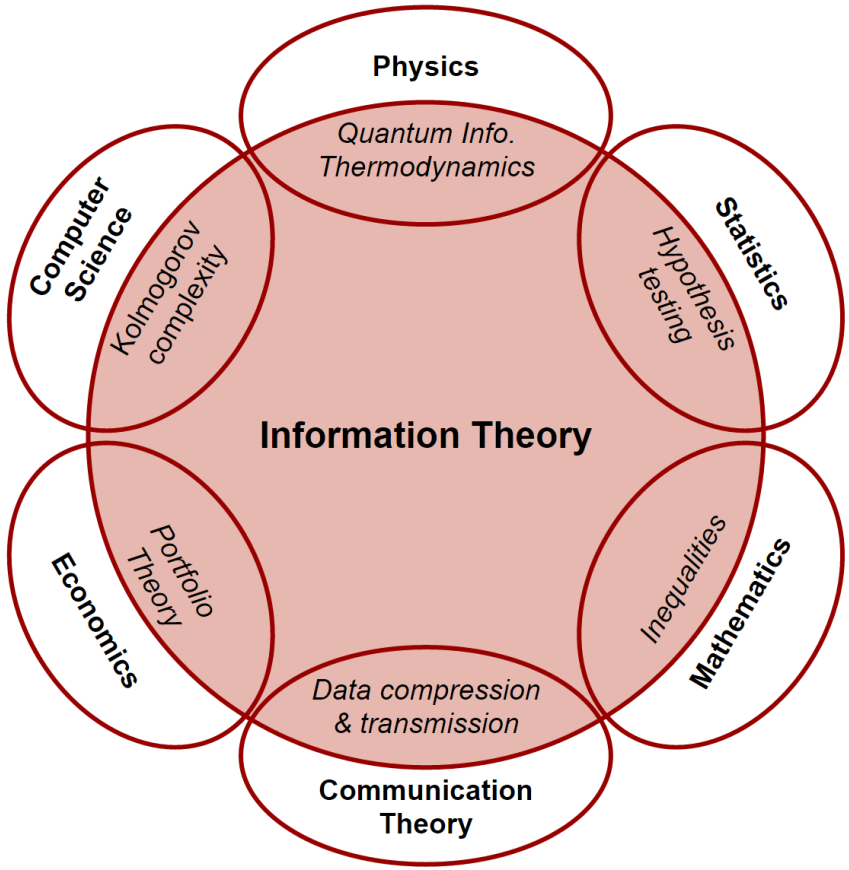
An introduction to Information Theory

- **Information theory** is the sub-area of computer science, mathematics, statistics, and engineering that deals with the definition and quantification of the concept of "information".
- Information theory was born in 1948, with the publication of Claude E. Shannon's seminal paper *"A Mathematical Theory of Communication"* [link].
- Shannon was interested in optimizing the use of communication channels (e.g., telephone and telegraph lines) for data transmission.

  By solving the problem of how to encode data for reliable and efficient data transmission over noisy (i.e., non-reliable) channels, Shannon pioneered the formalization of the mathematical concept of **information**.

An introduction to Information Theory

- Both the elegance and the fundamental aspect of Shannon's mathematical approach pushed information theory beyond data communication.
- Nowadays, information theory is an active area of research, both from the fundamental side and from the side of applications.
- New interpretations and applications of information theory arose in fields ranging from fundamental physicis to streaming videos efficiently on Netflix, from DNA sequencing to privacy on medical databases.



Overview of the course

Goals of this course

- This is an essentially theoretical course, in which we will cover the fundamentals of information theory.
- During the course we will use these fundamentals in numerous practical applications in areas including:

1. communication systems,
2. data transmission,
3. encoding of information,
4. error correction,
5. analysis of system complexity,
6. data compression,
7. hypothesis testing,
8. decision making,
9. security and privacy,
10. etc...

- That doesn't mean, however, that the applications of information theory are limited to these topics: they are just used as examples.

Goals of this course

- At the end of this course we should be able to answer the following questions:
  1. What is "*information*"? How can we *quantify* information?
  2. What does it mean to say one object *contains information* about another?
  3. How does data compression work?
  4. How do data-transmission channels work? How can we model computational systems as such channels?
  5. How to pick among concurrent hypotheses for a same phenomenon?
  6. What is "*randomness*"? How can we tell apart truly random phenomena from from phenomena that only look random?
  7. How to measure the *complexity of an object* (a string, a cup of tea, a puppy)?
  8. Are there information measures beyond Shannon entropy?
  9. How is information theory applied to fields such as security, privacy, and more?

Instructor

- Mário S. Alvim

  msalvim@dcc.ufmg.br

  `http://www.dcc.ufmg.br/~msalvim`

Important:
Any e-mails about this course should contain **[infotheory]** on its subject.

Bibliography

- **Text book:**
  - **Information Theory, Inference, and Learning Algorithms**
    David J. C. MacKay
    Cambridge University Press (2003)

    Available for download at [http://www.inference.phy.cam.ac.uk/mackay/itila] for on-screen viewing (printing not permitted, copyright restrictions apply as to a normal book).

- **Auxiliary book:**
  - **Elements of Information Theory**
    Thomas M. Cover, Joy A. Thomas
    Wiley-Interscience, 2nd Edition

Program

Part I - The Fundamentals

1. **Overview of the course and introduction**
   Slide-set L00 (MacKay, Chapter 1)

2. **Discrete Probability**
   Slide-set L01 (Rosen, Chapter 7, 7^{th} Ed.)

   - Saying goodbye to determinism.
   - Reviewing (or learning) how to reason in probabilistic terms.

3. **Probability, Entropy, and Inference / More About Inference**
   Slide-set L02 (MacKay, Chapter 2 / Chapter 3)

   - The different meanings of "probability".
   - Introduction to entropy and inference.

Program

Part II - Data Compression

4. **The Source Coding Theorem**
   Slide-set L03 (MacKay, Chapter 4)
   - How to measure the information content of a random variable.
   - Data compression: removing redundancy, keeping the essence.

5. **Symbol Codes**
   Slide-set L04 (MacKay, Chapter 5)
   - How much we can compress data?
   - Huffman codes.

6. **Stream Codes**
   Slide-set L05 (MacKay, Chapter 6)
   - Compressing on the fly.

Program

Part III - Channel Capacity

7. **Dependent Random Variables**
   Slide-set L06(MacKay, Chapter 8)

   - How much information one object carries about another.

8. **Communication Over a Noisy Channel**
   Slide-set L07 (MacKay, Chapter 9)

   - What is the capacity of a channel.
   - Calculating capacity in simple cases.

Program

Part IV - Probabilities and Inference

10. **Kolmogorov Complexity and Universal Probability**
Slide-set L08 (Cover & Thomas, Chapter 14)

    - Defining algorithmic (as opposed to probabilistic) entropy.
    - Describing the complexity of a cup of green tea, without sugar.
    - The *"full employment theorem"*.

09. **Decision Theory**
Slide-set L09 (MacKay, Chapter 36)

    - Making good decisions under uncertainty.
    - *"Cada escolha, uma renúncia, isto é a vida."* (Charlie Brown Jr.)

Program

Part V - Applications of information theory

10. **Advanced information measures**
    Slide-set L10

    - Beyond Shannon-entropy: *Rényi-entropy*, *Bayes vulnerability*, *Guessing entropy*.
    - The *g-vulnerability* framework.
    - An axiomatization of information measures.

11. **Information Theoretical Bases of Security and Privacy: SoS - Science of Security**
    Slide-set L11

    - Computational systems as channel, flow of information as data transmission.
    - The "*natural laws*" of security and privacy.

Grading

- Activities:
  - 3 exams: 60% of the final grade.
  - Seminar: 25% of the final grade.
  - Homework assignments: 15% of the final grade.

    (There will be about ten to twelve homework assignments, approximately one every week and a half. Keep up to date!)
- There will be a make-up exam, which:
  - replaces a missed exam,
  - will take place at the end of the semester, and
  - covers the whole program of the course.

Mário's Guiding principles

P1. **"Everything should be made as simple as possible, but no simpler."**

Each topic will be approached in a manner as clear as possible, but the intrinsic complexity of the topic won't be avoided. Challenges may lie ahead!

P2. **Good learning = mastering the maths + interpreting the results.**

We will develop both mathematical and pragmatic abilities in this course. Take advantage to improve what you need the most.

P3. **Intelligence is not synonymous with good results: reading the basic bibliography is essential, doing homework is essential.**

Understanding what is said in a lecture does not automatically mean you will score well in the exams. Reading and practicing are essential, and irreplaceable.

P4. **At this moment you have all the opportunities to suceed in this course, your success will depend on your attitude.**

It is your own individual effort that turns opportunities into results.

Are you interested in theoretical computer science?

- If have interest in working with, or maybe learn more about, topics like
  - **computational logic**,
  - **information theory**,
  - **decision and game theory**, or
  - **security and privacy**,

  come talk to me about opportunities regarding
  - undergraduate research projects ("iniciação científica"),
  - research projects, and
  - graduate programs (including internships abroad).