Homework no. 2

Information theory course AA. 2021/2022

A. SESSION GOALS

- Allow the students to familiarize with the definition, operational meaning and empirical measurements of information theoretic quantities for continuous variables.
- Allow the students to familiarize with the use of programming languages as tools to use for their investigations and the inference of observations from the obtained results.

B. SESSION TASKS

1. Using a programming language of your choice (e.g., Matlab, Python, etc.), design and implement a procedure that starting from an arbitrarily long list of independent realizations for a *continuous* random variable x provides an estimate of
   - the probability density function (pdf) of x, $p_x(a)$, $a \in \mathbb{R}$
   - the differential entropy of x, $h(x)$

   Note that your procedure should not assume any prior information on the statistics of x.

2. Test your procedure from 1. by generating a list of independent samples from the following distributions
   - a uniform distribution over [0, A]
   - a Gaussian distribution with mean $m_x = 0$ and variance $\sigma_x^2=p$
   - an exponential distribution with λ, i.e., $p_x(a) = \lambda e^{-\lambda a} 1(a)$

   Compare the empirical estimates of differential entropies to their theoretical values for different values of the distribution parameters A, p, λ and derive an empirical relationship between the number L of data values, the width Δ and the number N of pdf bins, the distribution parameters A, p, λ, and the precision of your estimate $\varepsilon = |\hat{h}(x)-h(x)|$.

3. Design and implement a procedure that starting from two arbitrarily long lists of realizations for two *continuous* random variables x and y provides estimates of
    - the pdf of x, $p_x(a)$, $a \in \mathbb{R}$, and y, $p_y(b)$, $b \in \mathbb{R}$, and their joint pdf, $p_{xy}(a,b)$, $(a,b) \in \mathbb{R}^2$
    - the joint and conditional differential entropies, $h(x,y)$ and $h(x|y)$, respectively
    - the Kullback-Leibler divergence, $D(p_x||p_y)$
    - the mutual information, $I(x;y)$
4. Test your procedure from 3. by generating a list of samples for x and y from the following distributions
    - x a uniform distribution in the interval [1, A], z a uniform distribution in the interval [-1, 1], and $y=x+z$
    - x a Gaussian distribution with mean m_x and variance σ_x^2, z a Gaussian distribution with mean m_z and variance σ_z^2 independent from x, and $y=ax+bz$

    Compare the empirical estimates to their theoretical values for different values of the distribution parameters A, m_x, m_z, σ_x^2, σ_z^2, and derive an empirical relationship between the number L of data values, the bin widths δ_x, δ_y for pdfs and the number of bins N_x, N_y, the distribution parameters A, m_x, m_z, σ_x^2, σ_z^2, and the precision of your estimates.

5. Write a report with the obtained results including figures and the conclusions that can be drawn. The report can be written using word or latex, as you prefer. You must deliver the final pdf. You need to attach also the code used for your evaluations.

The report must be delivered by May 23rd. You can upload it on the moodle of the course.

If you have any doubts about the assignment, we can meet on zoom on May 9th and May 18th from 9am to 10am
https://unipd.zoom.us/j/81689222190?pwd=R0dOQi84b2lZSEpWMEc4ZVRGRFVQUT09

In any case you can contact me at **guglielm@dei.unipd.it**