Homework no. 1

Information theory course AA. 2021/2022

A. SESSION GOALS

- Allow the students to familiarize with the definition, operational meaning and empirical measurements of basic information theoretical quantities.
- 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 discrete random variable *x* provides an estimate of
   - the alphabet of *x*, A_x
   - the probability mass distribution (pmd) of *x*, $p_x(a)$, $a \in A_x$
   - the Shannon entropy of *x*, $H(x)$
   - the collision entropy of *x*, $H_2(x)$
   - the guessing entropy of *x*, $H_{min}(x)$

   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 $A_x = \{1, \ldots, M\}$
   - a binary distribution with $p_x(0) = q$
   - a geometric distribution with λ, i.e., $p_x(k) = (1-\lambda)\lambda^k$

   Compare the empirical estimates of entropies to their theoretical values for different values of the distribution parameters *M, q, λ* and derive an empirical relationship between the number *L* of data values, the distribution parameters *M, q, λ*, and the relate precision of your estimate $\varepsilon = |\hat{H}(x)-H(x)|/H(x)$.

3. Design and implement a procedure that starting from two arbitrarily long lists of realizations for two discrete random variables x and y provides estimates of
    - the alphabet of x, A_x, and y, A_y
    - the pmd of x, $p_x(a)$, $a \in A_x$, and y, $p_y(b)$, $b \in A_y$, and their joint pmd, $p_{xy}(a,b)$
    - the joint and conditional 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, M], z a uniform distribution in the interval [-1, 1], and $y=x+z$
    - x a geometric distribution with parameter λ_x, i.e., $p_x(k) = (1- \lambda_x) \lambda_x^k$, and y a geometric distribution with parameter λ_y, i.e., $p_y(k) = (1- \lambda_y) \lambda_y^k$

    Compare the empirical estimates to their theoretical values for different values of the distribution parameters M, λ_x and λ_y and derive an empirical relationship between the number L of data values, the distribution parameters M, λ_x , λ_y and the relate 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 April 22nd. You can upload it on the moodle of the course.

If you have any doubts about the assignment, we can meet on zoom on April 6th and April 20th from 10am to 11am
https://unipd.zoom.us/j/88224218417?pwd=SitSUkduTDNDazlCTWlWYjh5dU1rdz09

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