Information Theory **DCC/ICEx/UFMG**
Prof. Mário S. Alvim **2024/1**

PROBLEM SET

Probability, Entropy, and Inference / The Source Coding Theorem
(MacKay - Chapter 2 / Chapter 4)

Necessary reading for this assignment:

- *Information Theory, Inference, and Learning Algorithms* (MacKay):

  Chapter 2

  – Chapter 2.4: *Definition of entropy and related functions*
  – Chapter 2.7: *Jensen's inequality for convex functions*

  Chapter 4

  – Chapter 4.1: *How to measure the information content of a random variable?*
  – Chapter 4.2: *Data compression*
  – Chapter 4.3: *Information content defined in terms of lossy compression*
  – Chapter 4.4: *Typicality*
  – Chapter 4.6: *Comments*

Note: The exercises are labeled according to their level of difficulty: `[Easy]`, `[Medium]` or `[Hard]`. This labeling, however, is subjective: different people may disagree on the perceived level of difficulty of any given exercise. Don't be discouraged when facing a hard exercise, you may find a solution that is simpler than the one the instructor had in mind!

Review questions.

1. Answer formally the following questions:

   (a) Define the Shannon information content $h(x)$ of the outcome x of a random experiment. Explain what the value $h(x)$ means.

   (b) Define the entropy $H(X)$ of an ensemble X. Explain what the value $H(X)$ means.

   (c) Define what is a convex $\smile$ function. Give at least two examples of functions that are convex $\smile$, and at least two of functions that are not.

   (d) State *Jensen's inequality.*

   (e) What is the formula for the raw bit content of an ensemble X? What does it mean?

   (f) Given an ensemble X, what is its smallest δ-sufficient subset S_δ?

   (g) Given an ensemble X and a value $0 < \delta < 1$, what is the essential bit content $H_\delta(X)$? What does it mean?

   (h) Shannon's source coding theorem can be stated as follows: If X is an ensemble with entropy $H(X) = H$ bits, then given any $\epsilon > 0$ and $0 < \delta < 1$, there exists a positive integer N_0 such that for $N > N_0$,

   $$\left| \frac{1}{N} H_\delta(X^N) - H \right| < \epsilon.$$

   Explain what it means for data compression.

Problems (Chapter 2).

2. (Lower bound for Shannon entropy) `[Easy]` Show that for every ensemble $X = (x, \mathcal{A}_X, \mathcal{P}_X)$, it is the case that $H(X) \geq 0$.

3. (Upper bound for Shannon entropy) The following exercises are designed so you can prove an upper bound for Shannon entropy.

   (a) (MacKay 2.21) `[Easy]`

   (b) (MacKay 2.22) `[Easy]`

   (c) (MacKay 2.25) `[Hard]` (Hint: Use Jensen's inequality!)

4. (Thomas&Cover 2.1) (The entropy of a countably infinite probability distribution) `[Medium]` A fair coin is flipped until the first head occurs. Let X denote the number of flips required. Find the entropy $H(X)$ in bits. (The following expressions may be useful: $\sum_{n=0}^{\infty} r^n = 1/(1-r)$, and $\sum_{n=0}^{\infty} nr^n = r/(1-r)^2$.)

Problems (Chapter 4).

5. (MacKay 4.2) `[Easy]`

6. (MacKay 4.5) `[Medium]`

7. (MacKay 4.9) `[Easy]`