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

PROBLEM SET

Advanced Information Measures
(Based on slide-set)

Necessary reading for this assignment:

- *Slide-set of the lecture on Advanced Information Measures*

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) Explain how probability distributions can be used to represent an agent's state of knowledge about the world.

   (b) What is a (prior) information measure? Explain what type of function it is (its domain and co-domain, and what it is supposed to mean).

   (c) Explain the essential components of a complete definition of an information measure: its mathematical definition, and operational significance.

   (d) Give the formal definition and operational significance of the following information measures.

      i. Shannon entropy.

      ii. Bayes vulnerability.

      iii. Guessing entropy.

   (e) Give the formal definition of a g-vulnerability, and explain how it can be used to model different operational scenarios.

   (f) What is the effect of a channel on the adversary's state of knowledge about a secret value? More precisely, how the posterior knowledge of an adversary can be represented, after they have observed the output of a channel?

   (g) How can the g-vulnerability framework be used to measure the amount of information contained in the adversary's posterior state of knowledge? More precisely, define the concept of posterior g-vulnerability.

   (h) How is the leakage of information of a channel, given a prior distribution on secrets, defined? What does it represent?

Exercises.

2. Consider an election in which k voters must choose among two candidates A and B, without the possibility of abstention or of null/blank votes. Naturally, each person's vote is sensitive information, so we can consider the sequence of all k votes to be a secret X to be protected.

   When the election is over,votes are computed and only the final tally is revealed publicly. (For instance, if there are $k = 8$ voters and the sequence of votes cast is $AABABAAB$, it will only be publicly revealed that the final tally is of 5 votes for candidate A and 3 votes for candidate B.)

(a) Assuming the prior distribution on secrets is the uniform prior π^u, what is the prior Bayes vulnerability of the secret?

(b) Represent the system of tallying for an election with $k = 3$ voters as an information-theoretic channel C with input X. Specify the output Y of the channel (i.e., the possible final tallies for the election), and draw the channel you obtain.

(c) Consider again an election with $k = 3$ voters. Assuming a uniform prior distribution π^u on secrets, compute the posterior Bayes vulnerability $V[\pi^u, C]$ of the election system. What is the resulting multiplicative Bayes leakage $\mathcal{L}(\pi, C) = {}^{V[\pi^u,C]}/_{V(\pi^u)}$, and what does this value mean?

(d) Here you will prove the following *"information flow law"*, which generalizes the result you computed in exercise (2c).

Theorem *Let π^u be a uniform prior over a set $\mathcal{X}$ of n elements. Let C be a deterministic channel (i.e., in which each input is mapped to exactly one output) with m possible output values (which impliest that $1 \leq m \leq n$). Then the resulting Bayes posterior vulnerability is given by*

$$V[\pi^u, C] = {}^{m}/_{n}.$$

(*Hint:* Notice that the deterministic channel C induces a partition on the secret set $\mathcal{X}$, in the sense that each output indicates that only a unique, mutually-exclusive, subset of secrets is possible. Apply this knowledge to the formula of $V[\pi^u, C]$ to derive the result.)

(e) Here you will go one step further in your search for *"information flow laws"*, proving the following corollary from what you proved in exercise (2d).

Corollary *Let π^u be a uniform prior over a set $\mathcal{X}$. Let C be a deterministic channel (i.e., in which each input is mapped to exactly one output) with m possible output values. Then the resulting multiplicative Bayes leakage is given by*

$$\frac{V[\pi^u, C]}{V[\pi^u]} = m.$$

(f) Use the *"information flow law"* from exercise (2e) to find the multiplicative Bayes leakage of the election system with $k = 3$ voters you computed in exercise (2c), assuming a uniform prior on secrets. Now compute the leakage for a system with a generic number k of voters, also assuming a uniform prior on secrets. Was it easier to just apply the law, or to do the whole set of calculations? This is the kind of law quantitative information flow is concerned with!