Quantitative Information Flow (QIF) **DCC/ICEx/UFMG**
Prof. Mário S. Alvim **2021/1**

SOLUTION OF PROBLEM SET

Introduction
(Chapter 01)

Necessary reading for this assignment:

- *The Science of Quantitative Information Flow* (Alvim, Chatzikokolakis, McIver, Morgan, Palamidessi, and Smith):
  - Chapter 1: *Introduction*
    - Chapter 1.1: *A first discussion of information leakage*
    - Chapter 1.2: *Looking ahead*

Review questions.

1. Explain what are the main goals of the study of quantitative information flow (QIF).

   Instructor's solution: Quantitative information flow is the area of knowledge concerned with quantifying how much sensitive information leaks through the observable behavior of a system.

2. Give an example of a system that could potentially leak sensitive information through observable outputs.

   Instructor's solution: A password checker may leak information about a secret password by accepting or rejecting a user's guess. If a correct guess is accepted, it's revealed that it corresponds to the correct password; if an incorrect guess is rejected, then it's revealed that the password is not that guess. In any case, the observable behavior of the password checker (either accepting or rejecting a guess) always leaks some information about the (secret) password value.

Exercises.

3. (Exercise 1.1) Recall dice channel C from §1.1, defined by $\mathsf{C}(r,w) = r{+}w$. Now consider a channel E that instead outputs the *maximum* of the two dice, so that $\mathsf{E}(r,w) = \max\{r,w\}$. Assuming a uniform prior distribution $\boldsymbol{\vartheta}$, find the additive and multiplicative Bayes leakage of E. What partition of $\mathcal{X}$ does E give?

   Instructor's solution: The set of possible outputs is $\{1,2,3,4,5,6\}$, so by Corollary 1.2 we have $\mathcal{L}_1^{\times}(\boldsymbol{\vartheta},\mathsf{E}) = 6$ and $\mathcal{L}_1^{+}(\boldsymbol{\vartheta},\mathsf{E}) = {}^5\!/_{36}$. The partition of $\mathcal{X}$ consists of 6 blocks:

$$
\begin{array}{l}
\{(1,1)\}\\
\{(1,2),(2,2),(2,1)\}\\
\{(1,3),(2,3),(3,3),(3,2),(3,1)\}\\
\{(1,4),(2,4),(3,4),(4,4),(4,3),(4,2),(4,1)\}\\
\{(1,5),(2,5),(3,5),(4,5),(5,5),(5,4),(5,3),(5,2),(5,1)\}\\
\{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6),(6,5),(6,4),(6,3),(6,2),(6,1)\}
\end{array}
$$

4. (Exercise 1.2) Consider an election in which k voters choose between candidates A and B. Ballots are supposed to be secret, of course, so we can take the sequence of votes cast to be the secret input X. (For example, if $k = 3$ then the set $\mathcal{X}$ of possible values for X is $\{AAA, AAB, ABA, ABB, BAA, BAB, BBA, BBB\}$.) Assuming a uniform prior $\boldsymbol{\vartheta}$, the prior Bayes vulnerability $V_1(\boldsymbol{\vartheta}) = 2^{-k}$, since there are 2^k possible sequences of votes, each occurring with probability 2^{-k}.

   When the election is tallied, the number of votes for each candidate is made public. (For example, when $k = 8$ we might get the vote sequence $AABABAAB$, whose tally is 5 votes for A and 3 votes for B.) Note that election tabulation can be seen as a *deterministic channel* T from X to Y, where Y is the tally of votes.

   (a) Given k, what is the multiplicative Bayes leakage $\mathcal{L}_1^{\times}(\boldsymbol{\vartheta}, \mathsf{T})$ of the election tabulation channel?

   (b) Suppose we want the *posterior Bayes vulnerability* $V_1[\boldsymbol{\vartheta} \triangleright \mathsf{T}]$ to be at most $1/8$. Determine the minimum value of k that achieves that bound.

Instructor's solution:

(a) The possible outputs of the election tabulation channel are all the possible tallies of votes; such a tally can be written as a pair (a, b) where a is the number of votes for A and b is the number of votes for B. Note that with k voters we always have $a + b = k$, since we are assuming no abstentions. So the number of possible tallies is $k + 1$, since the number of votes for A can be any number from 0 up to k. Hence by Corollary 1.2 we get $\mathcal{L}_1^{\times}(\boldsymbol{\vartheta}, \mathsf{T}) = k + 1$.

(b) Since $\boldsymbol{\vartheta}$ is uniform and T is deterministic, by Theorem 1.1 we have $V_1[\boldsymbol{\vartheta} \triangleright \mathsf{T}] = {(k+1)}/{2^k}$, since the number of possible channel outputs is $k + 1$ and the number of possible values of the secret is 2^k. Since 2^k grows much faster than $k + 1$, we see that $V_1[\boldsymbol{\vartheta} \triangleright \mathsf{T}]$ decreases as k grows:

k	1	2	3	4	5	6	7	8
$V_1[\boldsymbol{\vartheta} \triangleright \mathsf{T}]$	1	$3/4$	$1/2$	$5/16$	$3/16$	$7/64$	$1/16$	$9/256$

From the table, we see that if we want the posterior Bayes vulnerability to be at most $1/8$, then we need k to be at least 6.