DCC/ICEx/UFMG
Quantitative Information Flow
2021/1

Prof. Mário S. Alvim
First Exam
2021/07/01

Student:	**Registration number:**

Instructions:

i. This exam contains 2 page(s). Please verify that your copy of the exam is complete.

ii. **Write down your name on the first page of the exam**, and provide all of your answers **by hand**. You can use either **English or Portuguese**, as you prefer.

iii. You have **24 hours to complete the exam** and **submit your answers as a single scanned pdf file** to the proper interface on the course's Moodle. Make sure your submitted file is easily readable by a human.

iv. You are allowed to consult the course's textbook and slides, as well as your own notes, to complete this exam. **You are not allowed to consult the Internet, your colleagues, or any other people or sources to complete the exam**.

v. If you believe any question is under-specified, write down the assumptions you had to make to get to your answer and justify them as part of your answer to the question.

vi. Your answers will be evaluated in their clarity and conciseness. **Every answer must be thoroughly and properly justified.**

---

1. **(g-functions for all tastes – 6 marks)**

   Suppose you have a dataset X containing records of 1 000 distinct individuals born between 1 981 and 2 010 in the Southeastern region of Brazil. Each individual x's record consists in a tuple of four values, where

   - x[`name`] represents individual x's unique first name –that is, we assume there are no repeated first names in the dataset;
   - x[`state`] represents individual x's state of birth, which can be *ES*, *MG*, *RJ*, or *SP*;
   - x[`generation`] represents individual x's generation, which can be either *millenial* or *gen-Z*; and
   - x[`jeans`] represents individual x's preferred cut of jeans, which ban either be *skinny* or *baggy*.

   Assume additionally that the adversary knows the names of everyone in the dataset (but not necessarily any of the other fields), and that she wants to infer sensitive information about these individuals. Moreover, consider that gain functions must range in the interval [0,1].

   (a) Construct a set of actions and a gain function to capture the scenario in which the adversary knows an individual of name Mário is in the dataset X, and she benefits maximally from guessing his state of birth, and doesn't get any benefit from an incorrect guess.

   (b) Construct a set of actions and a gain function to capture the scenario in which the adversary benefits maximally by correctly naming any gen-Z'er in dataset X –no matter who they are– who wears skinny jeans; benefits only in 50% of the maximum by naming any millenial in dataset X –no matter who they are– who wears baggy jeans; and doesn't get any benefit from an incorrect guess.

2. **(It's six of one and half a dozen of the other – 6 marks)**

   Consider a secret set $\mathcal{X} = \{x_1, x_2\}$ and two alternative channels C and D below.

C	y_1	y_2	y_3
x_1	$2/15$	$1/3$	$8/15$
x_2	$1/15$	$2/3$	$4/15$

D	z_1	z_2	z_3
x_1	$2/3$	$2/15$	$1/5$
x_2	$1/3$	$4/15$	$2/5$

   Show that for any prior $\pi : \mathbb{D}\mathcal{X}$ and gain function $g : \mathbb{G}$, the additive and multiplicative leakage of both channels is the same.

3. **(Yet another question about biased coins – 10 marks)** Assume you have a biased coin with a bias of 70% towards one of the faces, but you don't know whether it's toward heads (H) or tails (T). Your goal is to find out what is the case for your coin.

   We can model this problem in QIF as follows. The secret set is $\mathcal{X} = \{0.3, 0.7\}$, where value 0.3 indicates that the coin yields heads 30% of the time and tails 70% of the time, and value 0.7 indicates that the coin yields heads 70% of the time, and tails 30% of the time.

   (a) If you don't have any hint on the bias of the coin, what should be your prior π on the secret set? What is the prior Bayes vulnerability $V_1(\pi)$ of the secret in this case?

   (b) To try and get some information on the kind of you coin you have in your hand, you can run the following experiment: flip the coin twice and observe the results. This experiment can be modeled as an appropriately chosen channel $\mathsf{C}\colon \mathcal{X} \to \mathbb{D}\mathcal{Y}$. Specify channel C's output set $\mathcal{Y}$, and draw its matrix.

   (c) Find the hyper distribution $[\pi \triangleright \mathsf{C}]$ resulting from pushing prior π through channel C. Explain what the inners and the outer in this hyper represent, in terms of the worlds in which the adversary may end up in after running the channel.

   (d) Compute the posterior Bayes vulnerability $V_1[\pi \triangleright \mathsf{C}]$ after your experiment. Give the corresponding values of the additive and multiplicative Bayes leakage, $\mathcal{L}_1^+(\pi, \mathsf{C})$ and $\mathcal{L}_1^\times(\pi, \mathsf{C})$, respectively, and explain what they mean.

4. **(There can be no leakage when the secret value is already known[1] – 3 marks**)

   Suppose that the prior π is a *point distribution*, i.e. that some x has probability 1 and all others have probability 0. Show that, regardless of the channel C and gain function g, there is no g-leakage from π as prior: that is, $\mathcal{L}_g^\times(\pi, C) = 1$ and $\mathcal{L}_g^+(\pi, C) = 0$. (Hint: just manipulate the definitions of prior- and posterior vulnerability to show that in this case $V_g[\pi \triangleright C] = V_g(\pi)$, and use this to immediately derive the desired conclusion.)

[1] I was in doubt about how to name this question. The other alternative was *"If you want to get 100% of the grade in an exam in a theoretical course, you need to be able to do this kind of proof"*. But in the end I decided that was too long and too explanatory...