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

SOLUTION OF PROBLEM SET

Posterior vulnerability and leakage
(Chapter 05)

Necessary reading for this assignment:

- *The Science of Quantitative Information Flow* (Alvim, Chatzikokolakis, McIver, Morgan, Palamidessi, and Smith):
  - Chapter 5: *Posterior vulnerability and leakage*
    - Chapter 5.1: *Posterior g-vulnerability and its basic properties*
    - Chapter 5.2: *Multiplicative and additive g-leakage*
    - Chapter 5.3: *A closer look at posterior Bayes vulnerability and Bayes leakage*
    - Chapter 5.4: *Measuring leakage with Shannon entropy*
    - Chapter 5.5: *More properties of posterior g-vulnerability and g-leakage*
    - Chapter 5.6: *Example channels and their leakage*
    - Chapter 5.7: *Max-case posterior g-vulnerability*

Review questions.

1. State the formulation of *posterior g-vulnerability* given by Definition 5.2, by Theorem 5.6, and by Theorem 5.7. Explain the purpose of each of these different (even if all equivalent) formulations.

   Instructor's solution: Given prior π, gain function $g\colon \mathbb{G}\mathcal{X}$ and channel C:

   - Definition 5.2 formulates posterior g-vulnerability $V_g[\pi \triangleright C]$ as the expected value of V_g over $[\pi \triangleright C]$, that is

   $$V_g[\pi \triangleright C] \;:=\; \sum_i a_i V_g(\delta^i)\;, \qquad \text{where} \qquad [\pi \triangleright C] \;=\; \sum_i a_i [\delta^i] \quad .$$

   This formulation is the very definition of posterior vulnerability, mapping a hyper-distribution representing the adversary's posterior knowledge directly to a value representing her information.

   - Theorem 5.6 formulates posterior g-vulnerability as

   $$V_g[\pi \triangleright \mathsf{C}] \quad = \quad \sum_{\substack{y \in \mathcal{Y} \\ p(y) \neq 0}} p(y) V_g(p_{X|y}) \quad .$$

   where C is a concrete channel implementing C.

   This formulation allows us to compute posterior g-vulnerability of a prior π, and concrete channel C directly from the posterior distributions $p_{X|y}$ for each output $y \in \mathcal{Y}$ of the channel, without having to compute the resulting hyper.

- Theorem 5.7 formulates posterior g-vulnerability as

$$V_g[\pi \triangleright \mathsf{C}] \quad = \quad \sum_{y \in \mathcal{Y}} \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \pi_x \mathsf{C}_{x,y} g(w,x) \quad ,$$

  where C is a concrete channel implementing C.
  This formulation allows us to compute posterior g-vulnerability directly from the prior π, the concrete channel C and the gain function g, without having to compute the resulting hyper or even the posterior distributions.

2. State the property of *monotonicity* relating prior- and posterior g-vulnerability. Explain informally what it means in terms of the information gained by an adversary by observing the behavior of a channel processing a secret value.

   Instructor's solution: The property of *monotonicity* states that posterior g-vulnerability is always greater than or equal to prior g-vulnerability: for any prior π, channel C and gain function $g\colon \mathbb{G}\mathcal{X}$, we have

$$V_g[\pi \triangleright C] \geq V_g(\pi) \ .$$

   This means that an adversary can never "lose information" (in average, since posterior g-vulnerability is defined as an average) by pushing a secret through a channel.

3. State the definition of *additive and multiplicative g-leakage.*

   Instructor's solution: Given prior distribution π, gain function $g\colon \mathbb{G}\mathcal{X}$, and channel C, the *multiplicative g-leakage* is

$$\mathcal{L}^{\times}_g(\pi, C) \ := \quad \frac{V_g[\pi \triangleright C]}{V_g(\pi)} \quad ,$$

   and the *additive g-leakage* is

$$\mathcal{L}^{+}_g(\pi, C) \ := \quad V_g[\pi \triangleright C] - V_g(\pi) \quad .$$

Exercises.

4. (Exercise 5.1) Prove Theorem 5.7.

   Instructor's solution: We have

$$\begin{array}{lll}
 & V_g[\pi \triangleright \mathsf{C}] & \\
= & \sum_{y|p(y)\neq 0} p(y) V_g(p_{X|y}) & \text{"Theorem 5.6"} \\
= & \sum_{y|p(y)\neq 0} p(y) \max_w \sum_x (p_{X|y})_x g(w,x) & \text{"Definition of } V_g\text{"} \\
= & \sum_{y|p(y)\neq 0} p(y) \max_w \sum_x p(x|y) g(w,x) & \text{"definition of } p_{X|y}\text{"} \\
= & \sum_{y|p(y)\neq 0} p(y) \max_w \sum_x \frac{p(x,y)}{p(y)} g(w,x) & \text{"definition of } p(x|y)\text{"} \\
= & \sum_{y|p(y)\neq 0} \max_w \sum_x p(x,y) g(w,x) & \text{"cancelation of } p(y)\text{"} \\
= & \sum_{y} \max_w \sum_x p(x,y) g(w,x) & \text{"}p(x,y) = 0 \text{ if } p(y) = 0\text{"} \\
= & \sum_{y} \max_w \sum_x \pi_x \mathsf{C}_{x,y} g(w,x) \quad . & \text{"definition of } p(x,y)\text{"}
\end{array}$$

Note that the normalizing factor $p(y)$ that converts from $p(x, y)$ to $p(x|y)$ is exactly *canceled out* by the weight $p(y)$ that is given to $V_g(p_{X|y})$.

5. (Exercise 5.3) It is noted before Definition 4.14 that *noninterference* is a traditional name for the "no leakage" property. Show that noninterference indeed implies "no g-leakage" — that is, show that if C satisfies noninterference, then for any prior π and gain function g, we have $\mathcal{L}_g^\times(\pi, C) = 1$ and $\mathcal{L}_g^+(\pi, C) = 0$.

   Instructor's solution: If C satisfies noninterference, than it maps any prior π to the point hyper-distribution $[\pi]$. Hence under Definition 5.29 we have $V_g[\pi \triangleright C] = V_g(\pi)$, implying that there is no g-leakage.

6. (Exercise 5.4) The *Monty Hall problem* is a famous brain teaser, based loosely on the old game show *Let's Make a Deal*, whose host was Monty Hall. In 1990, the problem appeared in the "Ask Marilyn" column of *Parade* magazine, formulated as follows:

   > Suppose you're on a game show, and you're given the choice of three doors: behind one door is a car; behind the other two are goats. You pick a door, say Door 1, and the host, who knows what's behind the doors, opens another door, say Door 3, which has a goat. He then says to you, "Do you want to pick Door 2 instead?"
   >
   > Is it to your advantage to switch your choice?

   This formulation is actually a bit imprecise about Monty Hall's behavior. What is intended is that Monty *always* opens a door that you did not choose and that contains a goat. (Since there are two doors containing goats, it is always possible for him to do that.) Also, he *always* gives you the option of switching. [1]

   To solve this puzzle, let us formulate Monty Hall as a probabilistic channel M (for "Monty") whose secret input X is the door containing the car, and whose output Y is the door that Monty opens after your initial choice, which we assume is Door 1. In the case when the car is behind Door 1, note that *both* Doors 2 and 3 contain goats, giving Monty a choice of which door to open. Here we assume that he makes this choice by flipping a fair coin, opening Door 2 if he gets *heads* and Door 3 if he gets *tails*. Hence the channel matrix is as follows:

M	2	3
1	$1/2$	$1/2$
2	0	1
3	1	0

   .

   Also, we assume a uniform prior $\boldsymbol{\vartheta} = (1/3, 1/3, 1/3)$, so that the car is equally likely to be behind each of the doors.

   (a) Calculate the hyper-distribution $[\boldsymbol{\vartheta} \triangleright \mathsf{M}]$.

   (b) Calculate the posterior Bayes vulnerability $V_1[\boldsymbol{\vartheta} \triangleright \mathsf{M}]$ and the multiplicative Bayes leakage $\mathcal{L}_1^\times(\boldsymbol{\vartheta}, \mathsf{M})$. Based on your calculations, should you stick with Door 1 or should you switch?

[1] A notable advantage of setting Quantitative Information Flow on a rigorous foundation is that it allows tricky problems (such as this one, Monty Hall) to be solved more or less *mechanically* — that is, without the need for deep thought. It calls to mind a wonderful quote from Alfred Whitehead:

> It is a profoundly erroneous truism, repeated by all copy-books and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle — they are strictly limited in number, they require fresh horses, and must only be made at decisive moments.

(c) Now assume that when the car is behind Door 1 (more generally, behind the door you choose), Monty uses a *biased* coin that gives *heads* with probability p and *tails* with probability $1-p$, for some p such that $0 \leq p \leq 1$. This changes the channel matrix to the following:

M	2	3
1	p	$1-p$
2	0	1
3	1	0

.

How does that change the results?

Instructor's solution:

(a) Given prior $\boldsymbol{\vartheta} = (1/3, 1/3, 1/3)$ and Monty Hall channel matrix

M	2	3
1	$1/2$	$1/2$
2	0	1
3	1	0

,

we get joint matrix

J	2	3
1	$1/6$	$1/6$
2	0	$1/3$
3	$1/3$	0

,

and marginal distribution $p_Y = (1/2, 1/2)$ and posterior distributions

	$p_{X\|2}$	$p_{X\|3}$
1	$1/3$	$1/3$
2	0	$2/3$
3	$2/3$	0

and finally hyper-distribution

$[\boldsymbol{\vartheta} \triangleright \mathsf{M}]$	$1/2$	$1/2$
1	$1/3$	$1/3$
2	0	$2/3$
3	$2/3$	0

.

(b) From Theorem 5.6 we can compute the posterior Bayes vulnerability by

$$V_1[\boldsymbol{\vartheta} \triangleright \mathsf{M}] = \sum_y p_Y(y) V_1(p_{X|y}) = 1/2 \cdot 2/3 + 1/2 \cdot 2/3 = 1/3 + 1/3 = 2/3 \quad .$$

On the other hand, by Theorem 5.15 we can just take the sum of the column maximums of J:

$$V_1[\boldsymbol{\vartheta} \triangleright \mathsf{M}] = \sum_y \max_x \mathsf{J}_{x,y} = 1/3 + 1/3 = 2/3 \quad ,$$

and indeed both techniques give the same answer.
Hence the multiplicative Bayes leakage is

$$\mathcal{L}_1^{\times}(\boldsymbol{\vartheta}, \mathsf{M}) \quad = \quad \frac{V_1[\boldsymbol{\vartheta} \triangleright \mathsf{M}]}{V_1(\boldsymbol{\vartheta})} \quad = \quad \frac{2/3}{1/3} \quad = \quad 2 \quad .$$

Or we can use Theorem 5.17 to get

$$\mathcal{L}_1^{\times}(\boldsymbol{\vartheta}, \mathsf{M}) \quad = \quad \sum_y \max_x \mathsf{M}_{x,y} \quad = \quad 1 + 1 \quad = \quad 2 \quad .$$

The posterior distributions make clear that the smart play is to *switch*. If Monty opens door 2, then $p_{X|2}$ shows that door 3 has probability $2/3$ of having the car, and if Monty opens door 3, then $p_{X|3}$ shows that door 2 has probability $2/3$ of having the car.

(c) If Monty's channel matrix is

M	2	3
1	p	$1-p$
2	0	1
3	1	0

then the hyper-distribution changes to

$[\vartheta \triangleright \mathsf{M}]$	$\frac{p+1}{3}$	$\frac{2-p}{3}$
1	$\frac{p}{p+1}$	$\frac{1-p}{2-p}$
2	0	$\frac{1}{2-p}$
3	$\frac{1}{p+1}$	0

.

Since $0 \leq p \leq 1$, we have $\frac{1}{p+1} \geq \frac{p}{p+1}$ and $\frac{1}{2-p} \geq \frac{1-p}{2-p}$, which means that switching is still the best strategy, no matter which door Monty opens.

But, interestingly, our success probability won't now be 2/3 regardless of the door Monty opens. Suppose $p = 0$, for instance. Then if Monty opens door 2, the car is *guaranteed* to be behind door 3. But if Monty opens door 3, the car is *equally likely* to be behind doors 1 and 2. Still, our *overall* success probability remains 2/3, regardless of p, since

$$V_1[\vartheta \triangleright \mathsf{M}] \;=\; \frac{p+1}{3} \cdot \frac{1}{p+1} \;+\; \frac{2-p}{3} \cdot \frac{1}{2-p} \;=\; \frac{1}{3} + \frac{1}{3} \;=\; \frac{2}{3} \quad .$$

This also means that the multiplicative Bayes leakage remains 2, regardless of p.

Also, note that that last fact follows immediately from Theorem 5.17, since the sum of the column maximums of M is 2, regardless of p.

7. (Exercise 5.7) Recall Exercise 4.2, which considers a password checker implementation with a timing side channel. Here we continue that topic in the more realistic setting of a 4-digit PIN ranging from 0000 to 9999. As before, an ideal password checker is modeled by a family of channels C^{g} with output set $\{\text{reject}, \text{accept}\}$. And a flawed implementation that compares the guess with the correct PIN digit by digit, rejecting as soon as a mismatch is found, is modeled by a family of channels D^{g} with output set $\{(\text{reject}, 1), (\text{reject}, 2), (\text{reject}, 3), (\text{reject}, 4), \text{accept}\}$.

   (a) Assuming a uniform prior $\vartheta = (1/10{,}000, 1/10{,}000, \ldots, 1/10{,}000)$, show that, for any guess g, the posterior Bayes vulnerabilities under C^{g} and D^{g} are $V_1[\vartheta \triangleright \mathsf{C}^{\mathsf{g}}] = 1/5000$ and $V_1[\vartheta \triangleright \mathsf{D}^{\mathsf{g}}] = 1/2000$.

   (b) Given that, one might be tempted to conclude that the difference in the posterior Bayes vulnerabilities is small enough that the side channel in D^{g} is not worth bothering about.

   But consider that in a typical PIN scenario, the adversary can enter a *sequence* of guesses $\mathsf{g}_1, \mathsf{g}_2, \ldots, \mathsf{g}_k$, thus running channels $\mathsf{D}^{\mathsf{g}_1}, \mathsf{D}^{\mathsf{g}_2}, \ldots, \mathsf{D}^{\mathsf{g}_k}$ for some moderate value of k. (Too many consecutive incorrect guesses might get her locked out.) Note that she can choose her guesses *adaptively*, which means that she can choose each guess g_i based on the outputs from the previous guesses $\mathsf{g}_1, \mathsf{g}_2, \ldots, \mathsf{g}_{i-1}$. How many adaptively chosen guesses to D^{g} does she need to determine the correct PIN in the worst case? In the average case?

   (c) In contrast, how many adaptively chosen guesses to the ideal password checker C^{g} does she need in the worst case? In the average case?

Instructor's solution:

   (a) Regardless of g, C^{g} and D^{g} have 2 and 5 possible outputs, respectively. The desired conclusions about posterior Bayes vulnerability then follow easily from Theorem 1.1 or Theorem 5.17.

(b) The key insight is that adaptively chosen guesses to D^{g} allow the adversary to determine the correct PIN digit by digit. She can first make guesses from the sequence `0000`, `1000`, `2000`, ..., `9000` until getting an output other than $(\text{reject}, 1)$. At that point she learns the first digit of the correct PIN (and more than that, as will be discussed in a moment). Thus she can learn the first digit in 10 guesses in the worst case, and in $^{1+2+\cdots+10}/_{10} = {}^{11}/_{2}$ guesses in the average case, since the first digit is uniformly distributed. Now suppose that the first guess giving an output other than $(\text{reject}, 1)$ is `6000`. Notice that this output also tells her whether the *second* digit is `0` or not: the output is $(\text{reject}, 2)$ iff the second digit is not `0`. Hence if the output is $(\text{reject}, 2)$, she can learn the second digit by making guesses from the sequence `6100`, `6200`, `6300`, ..., `6900`; otherwise she already knows that the second digit is `0`. Thus she can learn the second digit in 9 more guesses in the worst case, and in $^{0+1+\cdots+9}/_{10} = {}^{9}/_{2}$ more guesses in the average case. Continuing in this way, she can learn the correct PIN digit by digit in $10 + 9 + 9 + 9 = 37$ guesses in the worst case, and in $^{11}/_{2} + {}^{9}/_{2} + {}^{9}/_{2} + {}^{9}/_{2} = 19$ guesses in the average case. So ignoring the timing channel would be a huge mistake.

(c) The story for the ideal password checker C^{g} is completely different: any sequence $\mathsf{g}_1, \mathsf{g}_2, \ldots, \mathsf{g}_k$ of adaptively chosen guesses to C^{g} gives rise to a hyper-distribution with $k+1$ inners, k of which are point distributions, and one of which is a uniform distribution on all the unguessed PIN values. It follows that she needs 10000 guesses in the worst case, and $^{1+2+\cdots+10000}/_{10000} = {}^{10001}/_{2} = 5000.5$ guesses in the average case.