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

SOLUTION OF PROBLEM SET

Modeling secrets / On g-vulnerability
(Chapters 02 / 03)

Necessary reading for this assignment:

- *The Science of Quantitative Information Flow* (Alvim, Chatzikokolakis, McIver, Morgan, Palamidessi, and Smith):
  - Chapter 2: *Modeling secrets*
    - Chapter 2.1: *Secrets and probability distributions*
    - Chapter 2.2: *Shannon entropy*
    - Chapter 2.3: *Bayes vulnerability*
    - Chapter 2.4: *A more general view*
  - Chapter 3: *On g-vulnerability*
    - Chapter 3.1: *Basic definitions*
    - Chapter 3.2: *A catalog of gain functions*
    - Chapter 3.3: *Classes of gain functions*
    - Chapter 3.4: *Mathematical properties*
    - Chapter 3.5: *On "absolute" versus "relative" security*

Review questions.

1. Provide a brief description of the g-vulnerability framework, including a clear definition of secrets, actions, gain-functions, and g-vulnerability itself.

   Instructor's solution: Given a finite, nonempty set $\mathcal{X}$ (of possible secret values) and a nonempty set $\mathcal{W}$ (of possible actions), a gain function is a function $g\colon \mathcal{W}\times\mathcal{X}\rightarrow\mathbb{R}$. The value $g(w,x)$ specifies the gain that the adversary achieves by taking action w, when the value of the secret is x.

   The g-vulnerability of a prior distribution $\pi\colon\mathbb{D}\mathcal{X}$ on secrets in $\mathcal{X}$ is defined as

   $$V_g(\pi) \;:=\; \max_{w\in\mathcal{W}} \sum_{x\in\mathcal{X}} \pi_x\, g(w,x) \quad .$$

   (If $\mathcal{W}$ is infinite then the max should be replaced by sup.)

   The key idea is that when X has distribution π, a smart adversary should choose an action w that maximizes her expected gain $\sum_{x\in\mathcal{X}} \pi_x\, g(w,x)$ with respect to π.

2. Describe what is meant by the "operational interpretation/significance" of a vulnerability measure. How does the g-vulnerability framework allows for the expression of different operational interpretations?

Instructor's solution: The operational interpretation/significance of a vulnerability measure is a precise characterization of what the value returned by the measure means in a practical scenario.

In the g-vulnerability framework different operational interpretations for a vulnerability V_g can be achieved by changing the underlying gain function $g\colon \mathcal{W}\times\mathcal{X}\rightarrow\mathbb{R}$. Still, a gain function g is relevant to a particular operational scenario only if there really are actual actions corresponding to the elements of $\mathcal{W}$, whose effectiveness is correctly modeled by g.

Exercises.

3. (Exercise 2.1) Recall that 3 flips of a fair coin results in a uniform prior π^{coin} on the 8 values HHH, HHT, HTH, HTT, THH, THT, TTH, and TTT. What is the prior π^{bent} that results from 3 flips of a *bent* coin that gives heads with probability $^2/_3$ and tails with probability $^1/_3$? Compare the Shannon entropies $H(\pi^{coin})$ and $H(\pi^{bent})$ and the Bayes vulnerabilities $V_1(\pi^{coin})$ and $V_1(\pi^{bent})$.

Instructor's solution: We find that π^{bent} is

HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$^8/_{27}$	$^4/_{27}$	$^4/_{27}$	$^2/_{27}$	$^4/_{27}$	$^2/_{27}$	$^2/_{27}$	$^1/_{27}$

.

Hence we get $V_1(\pi^{coin}) = {}^1/_8$, $V_1(\pi^{bent}) = {}^8/_{27}$, $H(\pi^{coin}) = \log_2 8 = 3$, and

$$H(\pi^{bent}) \quad = \quad {}^8/_{27}\log_2 {}^{27}/_8 + 3\cdot{}^4/_{27}\log_2 {}^{27}/_4 + 3\cdot{}^2/_{27}\log_2 {}^{27}/_2 + {}^1/_{27}\log_2 27 \quad \approx \quad 2.75489 \quad .$$

4. (Exercise 3.1) Let us explore g-vulnerability in the context of a hypothetical lottery. Suppose that the lottery sells tickets for \$2 each, in which the purchaser marks 6 choices out of the numbers from 1 to 40. (For example, the purchaser might mark 3, 23, 24, 31, 33, and 40.)

   Then a drawing is held in which 6 such numbers are selected randomly. Suppose that the rules are that a ticket wins only if it exactly matches the 6 selected numbers, in which case the player gets \$5 million; otherwise, the player gets nothing. (Real lotteries have more complicated rules!)

   (a) Design a gain function g suitable for modeling this lottery. The set $\mathcal{X}$ of possible secret values x is the set of all sets of 6 numbers from 1 to 40. Let the set $\mathcal{W}$ of possible actions include "buy t" for each set t of 6 numbers from 1 to 40, along with the action "don't play".

   (b) Calculate the g-vulnerability $V_g(\boldsymbol{\vartheta})$, assuming that $\boldsymbol{\vartheta}$ is uniform. Which action is optimal?

Instructor's solution:

(a) A suitable gain function g is as follows.

$$g(w,x) = \begin{cases} 4999998, & \text{if } w \text{ is "buy } t\text{", where } t = x \\ -2, & \text{if } w \text{ is "buy t", where } t \neq x \\ 0, & \text{if } w \text{ is "don't play"} \end{cases}$$

Note that the gain for buying a winning lottery ticket is not quite \$5 million, since the cost of buying the ticket (\$2) needs to be deducted.

(b) First we need to figure out $\boldsymbol{\vartheta}$ by counting the number of possible lottery tickets. Since we are selecting 6 numbers out of 40, the number is given by the binomial coefficient

$$\binom{40}{6} = \frac{40!}{34!\;6!} = \frac{40\cdot 39\cdot 38\cdot 37\cdot 36\cdot 35}{6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1} = 3838380$$

(This is because there are 40 choices for the first number, 39 for the second, 38 for the third, 37 for the fourth, 36 for the fifth, and 35 for the sixth; but each such sequence of six distinct numbers can be ordered in 6! ways, and order is irrelevant for the lottery.) So, since we are assuming a uniform prior distribution $\boldsymbol{\vartheta}$, we have $\boldsymbol{\vartheta}_x = {}^1\!/_{3838380}$ for every x.

Now we calculate the prior g-vulnerability based on its definition:

$$V_g(\boldsymbol{\vartheta}) = \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \boldsymbol{\vartheta}_x \, g(w,x) \quad .$$

If w is "don't play", then we see that $g(w,x) = 0$ for every x, and hence $\sum_x \boldsymbol{\vartheta}_x \, g(w,x) = 0$.
If w is "buy t", then we see that $g(w,x) = 4999998$ for $x = t$ and $g(w,x) = -2$ for every other x. Hence

$$\sum_x \boldsymbol{\vartheta}_x \, g(w,x) \quad = \quad \frac{1}{3838380} \cdot 4999998 + \frac{3838379}{3838380} \cdot (-2) \quad = \quad -\frac{133838}{191919} \quad \approx \quad -0.697 \quad ,$$

so the expected *loss* from action "buy t" is about 70 cents, regardless of t. Hence the prior g-vulnerability $V_g(\boldsymbol{\vartheta}) = 0$ and the optimal action is "don't play".

5. (Exercise 3.2) Let g be a gain function with the following matrix representation:

G	x_1	x_2
w_1	3	-1
w_2	-8	2

Give a prior π such that $V_g(\pi) < 0$, which implies that $g \notin \mathbb{G}\mathcal{X}$.

Instructor's solution: If the prior $\pi = (p, 1-p)$, then the expected gain for w_1 is $3p + (-1)(1-p) = 4p - 1$, which is negative if $p < {}^1\!/_4$. And the expected gain for w_2 is $(-8)p + 2(1-p) = 2 - 10p$, which is negative if $p > {}^1\!/_5$. Hence $V_g(\pi) < 0$ iff ${}^1\!/_5 < p < {}^1\!/_4$.