

INFORMATION THEORY

University of Amsterdam, 2019

TEACHERS: Christian Schaffner, Yfke Dulek TAs: Esteban Landerreche, Maximilian Siemers

TEAM NAME: Hadamard

STUDENT NAMES: Leonardo Romor, Maurice Frank, Simone Astarita, XinYu Fu, Yije Zhang

Homework set 1

Contents

| | |
|-----------|---|
| Problem 1 | 2 |
| a. | 2 |
| b. | 2 |
| c. | 2 |
| Problem 2 | 4 |
| a. | 4 |
| b. | 4 |
| c. | 4 |
| d. | 5 |
| Problem 3 | 6 |
| Problem 4 | 7 |
| a. | 7 |
| b. | 8 |
| c. | 9 |
| d. | 9 |
| e. | 9 |

Problem 1 (7 pt): Two coins and a die

You have two (fair) coins and a (fair) 4-sided die with outcomes $\{1, 2, 3, 4\}$. Let X be the number of heads after flipping the two coins and let Y be the result of rolling the die. Let Z be the average of X and Y .

(2 pt) **a.** What is the distribution P_Z of Z ?

Solution: We can first write down P_X in Table 1 and P_Y in Table 2. Then, it is easy to get the distribution of Z by simple counting. P_Z is shown in Table 3.

Table 1: Distribution of X

| X | 0 | 1 | 2 |
|--------|---------------|---------------|---------------|
| $P(X)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

Table 2: Distribution of Y

| Y | 1 | 2 | 3 | 4 |
|--------|---------------|---------------|---------------|---------------|
| $P(Y)$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

Table 3: Distribution of Z

| Z | $\frac{1}{2}$ | 1 | $\frac{3}{2}$ | 2 | $\frac{5}{2}$ | 3 |
|--------|----------------|----------------|---------------|---------------|----------------|----------------|
| $P(Z)$ | $\frac{1}{16}$ | $\frac{3}{16}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{3}{16}$ | $\frac{1}{16}$ |

□

(3 pt) **b.** Compute the variances of X , Y and Z .

Solution: Just apply the definition of variance, easy to get:

$$\mathbb{E}(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1 \quad (1)$$

$$\text{Var}(X) = (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (2-1)^2 \times \frac{1}{4} = \frac{1}{2} \quad (2)$$

$$\mathbb{E}(Y) = 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} + 4 \times \frac{1}{4} = \frac{5}{2} \quad (3)$$

$$\text{Var}(Y) = (1-\frac{5}{2})^2 \times \frac{1}{4} + (2-\frac{5}{2})^2 \times \frac{1}{4} + (3-\frac{5}{2})^2 \times \frac{1}{4} + (4-\frac{5}{2})^2 \times \frac{1}{4} = \frac{5}{4} \quad (4)$$

$$\mathbb{E}(Z) = \frac{1}{2} \times \frac{1}{16} + 1 \times \frac{3}{16} + \frac{3}{2} \times \frac{1}{4} + 2 \times \frac{1}{4} + \frac{5}{2} \times \frac{3}{16} + 3 \times \frac{1}{16} = \frac{7}{4} \quad (5)$$

$$\text{Var}(Z) = (\frac{1}{2} - \frac{7}{4})^2 \times \frac{1}{16} + (1 - \frac{7}{4})^2 \times \frac{3}{16} \quad (6)$$

$$+ (\frac{3}{2} - \frac{7}{4})^2 \times \frac{1}{4} + (2 - \frac{7}{4})^2 \times \frac{1}{4} + (\frac{5}{2} - \frac{7}{4})^2 \times \frac{3}{16} + (3 - \frac{7}{4})^2 \times \frac{1}{16} = \frac{7}{16} \quad (7)$$

□

(2 pt) **c.** You play the following game. If $2X \geq Y$, you win X^2 euros and otherwise you lose 1 euro. What is your expected total gain or loss after playing this game 40 times?

Solution: Let's define a new random variable $A = 2X - Y$ for one round. Also, define the return function $f(A)$ on A as:

$$f(a) = \begin{cases} -1 & a < 0 \\ X^2 & 0 \leq a \end{cases} \quad (8)$$

According to the question, the expectation $\mathbb{E}[f(A)]$ can be computed as:

$$\mathbb{E}[f(A)] = \sum_a p(a)f(a) = P(A \geq 0) \cdot X^2 + P(A < 0) \cdot (-1) \quad (9)$$

where we have $P(A \geq 0)$ as:

$$P(A \geq 0) = P(X = 1) \cdot P(Y = 1 \text{ or } 2) + P(X = 2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \quad (10)$$

and $P(A < 0)$ as:

$$P(A < 0) = 1 - P(A \geq 0) = \frac{1}{2} \quad (11)$$

so come back to the $\mathbb{E}[f(A)]$:

$$\mathbb{E}[f(A)] = P(X = 1, Y = 1 \text{ or } 2) \cdot 1^2 + P(X = 2) \cdot (2)^2 + P(A < 0) \cdot (-1) = \frac{3}{4} \quad (12)$$

Since in each round, the flipping of a dice and a coin is clearly independent of the flipping in last round, $\{A_i\}_1^{40}$ are i.i.d. random variables. Hence, the expected total gain or loss after playing this game 40 times is:

$$\mathbb{E}\left[\sum_{i=1}^{40} f(A_i)\right] = \mathbb{E}[40 \cdot f(A)] = 40 \cdot \mathbb{E}[f(A)] = 40 \cdot \frac{3}{4} = 30 \quad (13)$$

□

Problem 2 (9 pt): Deriving the weak law of large numbers

- (2 pt) **a. (Markov's inequality)** For any real non-negative random variable X and any $t > 0$, show that

$$P[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Solution:

Given indicator function:

$$tI_{x \geq t} \leq X$$

when $x < t$, $tI_{x \geq t} = 0 \leq X$, else when $x \geq t$, $tI_{x \geq t} = t \leq X$

Get the expected value for left and right sides, we obtain:

$$t \mathbb{E}(I_{x \geq t}) \leq \mathbb{E}(X)$$

Substitute $\mathbb{E}(I_{x \geq t}) = P[X \geq t]$ and after items rearrangement, we obtain:

$$P[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

□

- (1 pt) **b.** Exhibit a random variable (which can depend on t) that achieves this inequality with equality.

Solution: Either $x = 0$ or $t = 0$

□

- (3 pt) **c. (Chebyshev's inequality)** Let Y be a random variable with mean μ and variance σ^2 . Show that for any $\varepsilon > 0$,

$$P[|Y - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}.$$

Hint: Define a random variable $X := (Y - \mu)^2$.

Solution: We define a random variable $X := (Y - \mu)^2$. Therefore:

$$\begin{aligned} P[|Y - \mu| \geq \varepsilon] &= P[(Y - \mu)^2 \geq \varepsilon^2] \\ &= P[X \geq \varepsilon^2] \\ \sigma^2 &= \mathbb{E}[(y - \mu)^2] \\ &= \mathbb{E}[X] \end{aligned}$$

Using now the Markov inequality with $t = \varepsilon^2$ we have:

$$P[X \geq \varepsilon^2] \leq \frac{\mathbb{E}[X]}{\varepsilon^2}$$

which proves the Chebyshev inequality.

□

(3 pt)

d. (The weak law of large numbers.) Let Z_1, Z_2, \dots, Z_n real i.i.d. random variables with mean $\mu = \mathbb{E}[Z_i]$ and variance $\sigma^2 = \mathbb{E}[(Z_i - \mu)^2] < \infty$. Define the random variables $S_n = \frac{1}{n} \sum_{i=1}^n Z_i$. Show that

$$P[|S_n - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Thus, $P[|S_n - \mu| \geq \varepsilon] \rightarrow 0$ as $n \rightarrow \infty$. This is known as the weak law of large numbers (which we will use heavily in Week 03).

Solution:

Given Chebyshev's inequality

$$P[|S_n - \mu| \geq \varepsilon] \leq \frac{\text{Var}[S_n]}{\varepsilon^2}.$$

Substitute

$$\text{Var}[S_n] = \frac{\sum_{i=1}^n \text{Var}[Z_i]}{n^2} = \frac{\sigma^2}{n}$$

We obtain the weak law of large numbers

$$P[|S_n - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2}.$$

where $P[|S_n - \mu| \geq \varepsilon] \rightarrow 0$ as $n \rightarrow \infty$

□

Problem 3 (3 pt): Multiple-choice test

A multiple-choice exam has 4 choices for each question. A student has studied enough so that the probability she will know the answer to a question is 0.5, the probability that she will be able to eliminate one choice is 0.25, otherwise all 4 choices seem equally plausible. If she knows the answer she will get the question right. If not she has to guess from the 3 or 4 choices.

As the teacher you want the test to measure what the student knows. If the student answers a question correctly, what is the probability she knew the answer? Give your answer with three decimals of precision.

Solution: First we look at the probability of guessing a MC-question correctly. For a question with n answers we need the possible combinations of correct answers (*multiple*):

$$\sum_{i=1}^n \binom{n}{i} = 2^n - 1$$

Making the probability of guessing correctly the reciprocal of that.

There are three paths to the correct solution: she knows the solution ($\frac{1}{2}$), she eliminated one solution and guessed ($\frac{1}{4} \cdot \frac{1}{2^3-1}$) or she completely guessed ($\frac{1}{4} \cdot \frac{1}{2^4-1}$).

Therefore the probability of she not guessing is:

$$p(\text{knowledge} = \text{yes} | \text{answer} = \text{correct}) = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{28} + \frac{1}{60}} = \frac{105}{116} = 0.905 = 90.517\%$$

□

Problem 4 (8 pt): Computing Variational Distance (programming)

The *total variation distance* between two probability distributions P and Q over the finite alphabet \mathcal{X} is defined as

$$\|P - Q\| := \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|$$

This distance measure is symmetric, fulfills the triangle inequality and is normalized, i.e. it is 0 iff $P = Q$ and 1 iff P and Q have disjoint support.

The *collision probability* of a distribution P over finite alphabet \mathcal{X} is defined as

$$\text{Coll}(P) := \sum_{x \in \mathcal{X}} P(x)^2$$

In this exercise, we are going to analyze the letter frequencies of *Alice in Wonderland* in five different languages: English, German, Esperanto, Italian and Finnish. You can find all necessary files here: <https://github.com/cschaffner/InformationTheory/tree/master/Problems/HW1>. Hereby, we are going to consider only the 26 English letters (without space) and ignore that languages like German and Finnish have important other letters such as ä, ö, ü.

For $\text{lang} \in \{\text{eng, ger, esp, ita, fin}\}$, let P_{lang} be the frequency distribution of the 26 English letters (without space) of *Alice in Wonderland*.

- (2 pt) **a.** Compute all pairwise variational distances $\|P_{\text{lang}} - P_{\text{lang}'}\|$ for $\text{lang} \neq \text{lang}' \in \{\text{eng, ger, esp, ita, fin}\}$. Which two languages are closest, which two are furthest apart in terms of variational distance?

Note: for this exercise and any future programming exercises, you do not have to submit your code. In the pdf that you hand in, describe in a few sentences which general strategy you used (e.g., what quantities did you compute and in what order?), any choices you made (e.g., how did you treat uppercase letters? How did you deal with edge cases?), and any ‘sanity check’ computations you may have performed (e.g., did you check what the variational distance between a text file and itself was?) By adding this information, you may still receive partial credit for your approach, even if your final numerical answer is incorrect.

Solution:

To evaluate the *variational distance* we performed some pre-processing on the provided data. First, we converted every uppercase letter in lowercase. A second pass was done to remove any punctuation or special non ascii-lowercase characters (26 characters “abcdefghijklmnopqrstuvwxyz”). This was done by providing a translation table. After preparing the inputs we computed the frequencies and relative frequencies treating them as the empirical probability distributions of the alphabet.

In [Figure 1](#) the relative frequencies are shown.

Novel alphabet frequencies by language and for the permuted cipher.

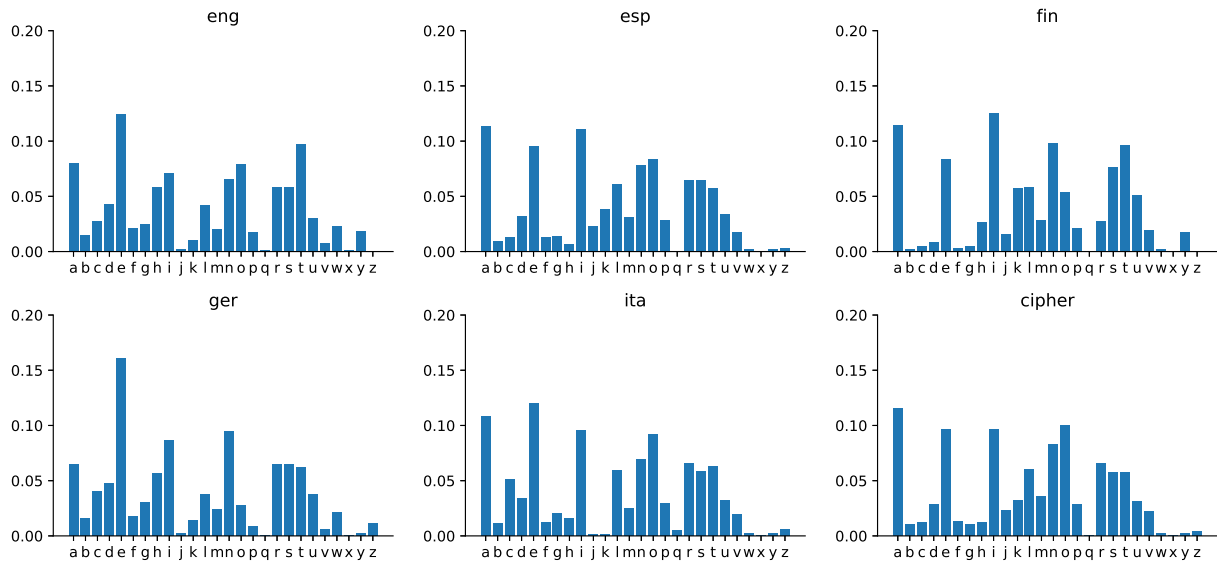


Figure 1: Relative frequencies of the 6 provided inputs.

To sanity check, we tested the properties of such distance, such as identity of indiscernibles, symmetry and subadditivity.

We then computed all the possible combinations resulting in table Table 4:

| P \ Q | eng | esp | fin | ger | ita | cipher |
|--------|------|------|------|------|------|--------|
| eng | 0 | 0.21 | 0.26 | 0.14 | 0.16 | 0.21 |
| esp | 0.21 | 0 | 0.16 | 0.24 | 0.11 | 0.04 |
| fin | 0.26 | 0.16 | 0 | 0.3 | 0.24 | 0.18 |
| ger | 0.14 | 0.24 | 0.3 | 0 | 0.19 | 0.24 |
| ita | 0.16 | 0.11 | 0.24 | 0.19 | 0 | 0.21 |
| cipher | 0.21 | 0.04 | 0.18 | 0.24 | 0.21 | 0 |

Table 4: Pairwise variational distance.

If we exclude the cipher from our results, we have that the highest variational distance is between *fin-ger* with a distance of 0.3 while the lowest is between *esp-ita* with a distance of 0.11.

□

- (2 pt) **b.** Compute the five collision probabilities $Coll(P_{lang})$ for $lang \in \{eng, ger, esp, ita, fin\}$.
Note: You do not have to submit your code.

Solution:

| l | fin | ger | ita | cipher | esp | eng |
|-----------|-------|-------|-------|--------|-------|-------|
| $Call(P)$ | 0.076 | 0.072 | 0.071 | 0.070 | 0.069 | 0.065 |

Table 5: Collision probabilities.

The collision probabilities were calculated from the previous relative frequencies.

☐

- (1 pt) c. Why is it called collision probability?

Solution: It's called collision probability because it represents the probability that two independent samples of random variables X_1, X_2 being sampled from the same distribution have of resulting in equal values.

☐

- (2 pt) d. You are [given the file](#) `permuted_cipher.txt` that has been encrypted by (first removing spaces and then) shuffling around the characters (i.e. by applying a permutation cipher). Note that this kind of encryption preserves the letter frequencies. Compute the frequency distribution P_{cipher} and figure out which language the original text was by picking the one that minimizes by the variational distance $\|P_{cipher} - P_{lang}\|$ with $lang \in \{eng, ger, esp, ita, fin\}$ as above.

Note: You do not have to submit your code.

Solution: The language that seems to minimize the variational distance with the cipher is *esp* as shown in [Table 4](#).

☐

- (1 pt) e. Would you have picked the same language when comparing the collision probability $Coll(P_{cipher})$ to the ones above?

Solution: Yes, as the distribution having the most similar collision probability with *cipher* is indeed *esp*.

☐