TEACHER: Christian Schaffner, TAs: Yfke Dulek, Esteban Landerreche, Kyah Smaal

Team homework set 1

Problem 1: Deriving the weak law of large numbers (9pt)

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

$$P_X(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

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

(c) **(3pt)** (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 $X := (Y - \mu)^2$.

(d) **(3pt)** (The weak law of large numbers.) Let $Z_1, Z_2, ..., Z_n$ real i.i.d. random variables with mean $\mu = \mathbb{E}[Z_i]$ and variance $\sigma^2 = \mathbb{E}[(X_i - \mu)^2] < \infty$. 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, $P(|S_n - \mu| \geq \varepsilon) \to 0$ as $n \to \infty$. This is known as the weak law of large numbers (which we will use heavily in Week 03).

Problem 2: Birthday party (9pt)

20 ILLC members are having a party.

(a) **(3pts)** To prepare, they need to choose 3 people to set the table, 2 people to bake cake and 6 people to clean up. Each person can only do 1 task (this doesn't add up to 20, the rest of the people don't help). In how many different ways can they choose which people perform these tasks?

(b) **(3pts)** The party crowd consists of 5 staff members and 15 students, and tea and coffee is served after the cake. It turns out that all staff members don't like tea. If they only give tea to 10 of the 20 people, what is the probability that only students get tea?

(c) **(3pts)** If they only give tea to 10 of the 20 people, what is the probability that 9 students and 1 staff member gets tea?

Problem 3: Expectation and variance (4pt)

(a) **(2pt)** Show that expectation is linear: for arbitrary $a, b \in \mathbb{R}$ and arbitrary real random variables X, Y,

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

(b) **(2pt)** Use Jensen's inequality to show that $\mathrm{Var}(X) \geq 0$ for any X.

Problem 4: Two coins and a die (8pt)

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.

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

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

(c) **(3pt)** 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?

Problem 5: Fermi estimates (8pt)

Read up on what a Fermi estimate is. Give a Fermi estimate for the *daily number of git commits in the world*.

In this exercise, it is more important to be clear about your arguments and assumptions rather than getting the correct result.

(a) **(4pt)** Give your Fermi estimate without looking up any numbers on the internet.

(b) **(4pt)** Look up the number for your estimates from the previous sub-exercise. Possibly revise your strategy to do the estimate. Compare the new final outcome to what you got previously as endresult in (a).

One team member should post your answers to this discussion forum, don't forget to state your team name.

Problem 6: Programming: Computing Variational Distance (8pt)

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

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

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

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

(b) **(2pt)** Compute the five collision probabilities $Coll(P_{lang})$ for $lang \in \{\text{eng, ger, esp, ita, fin}\}$.

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

(d) **(2pt)** 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 $lang \in \{\text{eng, ger, esp, ita, fin}\}$ as above.

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