

Homework 1: April 2021

Instructor: Shashank Vatedka

Instructions: You are encouraged to discuss and collaborate with your classmates. However, you must explicitly mention at the top of your submission who you collaborated with. Copying is NOT permitted, and solutions must be written independently and in your own words.

Homeworks must be submitted on Google classroom. Please scan a copy of your handwritten assignment and upload as pdf with filename `<your ID>_HW<homework no>.pdf`. Example: `EEB19BTECH00000_HW1.pdf`.

For programming questions, submit as separate files. Please use the naming convention `<your ID>_HW<homework no>_problem<problem no>.*`. Example: `EEB19BTECH00000_HW1_problem1.c`

Exercise 1.1 (35 points). Suppose that we are given n iid samples $\underline{X} = X_1, X_2, \dots, X_n$ from an unknown distribution p_X over $\{1, 2, \dots, m\}$. We know m , but have to estimate p_X .

- Derive the maximum likelihood estimator for p_X . Is it unbiased?
- Derive the moment matching estimator for p_X . Is it unbiased? Can you comment on the relative performance of this and the ML estimator?
- Consider a third estimator (let us call this the add-1 estimator) for p_X : For $a = 1, 2, \dots, m$, and an observation \underline{x} set

$$\hat{p}_X(a) = \frac{n_a(\underline{x}) + 1}{n + m},$$

where $n_a(\underline{x})$ is the number of occurrences of symbol a in \underline{x} . Is this unbiased? Is it asymptotically unbiased?

Exercise 1.2 (30 points). You are given a file containing symbols from $1, 2, \dots, 7$. Compute the ML and the add-1 estimates for the pmf. Also for each estimate \hat{p}_X compute the mean squared error

$$MSE = \sum_{x \in \mathcal{X}} (p_X(x) - \hat{p}_X(x))^2$$

and the total variation distance

$$TV(p_X, \hat{p}_X) = 0.5 \times \sum_{x \in \mathcal{X}} |p_X(x) - \hat{p}_X(x)|$$

Use these to estimate the entropy: find $H(\hat{p}_X)$ for each of the estimators, and compare $|H(\hat{p}_X) - H(p_X)|$.

Exercise 1.3 (35 points). A major component of any communication system is estimating the channel. Consider the following setup where if we transmit a sequence of symbols x_0, x_1, \dots, x_{n-1} , the receiver obtains

$$y_i = \begin{cases} h_0 x_0 + h_1 x_{n-1} + z_0 & \text{if } i = 0 \\ h_0 x_i + h_1 x_{i-1} + z_i & \text{if } i = 2, 3, \dots, n-1 \end{cases}$$

Here, the receiver knows x_0, \dots, x_{n-1} (these are *pilot symbols* that are decided beforehand), and wants to estimate (h_0, h_1) from y_0, y_1, \dots, y_{n-1} . Also, z_0, z_1, \dots, z_{n-1} are iid Gaussian random variables with mean zero and variance σ^2 . We assume that σ^2 is also known.

Derive the maximum likelihood estimator for h_0, h_1 .

Hint: It might help to write the above as a matrix equation.

You are also given a template for a python program `template.py` to simulate channel estimation (if you want to use a different programming language, please convert it line-by-line to the language of your choice). Use this to simulate and plot the MSE for different σ^2 . You may only make changes to the function `estimator(x,y)`. Leave the rest as it is.