Freie Universität Berlin
Liu-Wei Wang, Max von Kleist

8. Homework
Foundations of Mathematics and Statistics

Deadline: December 18, 10:00 (**before** the lecture)

The homework should be worked out in groups. Pen & paper exercises will be discussed on the board. Python programs must be submitted via Whiteboard, plots printed and handed in (write the names of the group members and their student numbers on the sheet).

Homework 1 (Likelihood - univariate, (programming counts 1/6 + 1/6))
You observed the following sequence of coin flips (y_i) = ('HHTHHTTH'). You do not know whether the coin is fair and want to determine the value $\theta = p(y_i =' H')$ from the data.
a) [pen & paper] Write down the probability density function for each possible measurement, then write down the **log-likelihood** of your observations (= sequence of coin flips), $\log\left(\mathcal{L}_{(y_i)|\theta}\right)$.
b) [pen & paper] Derive the maximum likelihood estimate θ^* (via pen & paper).
c) [Programming, submit via whiteboard] Write a program that computes the likelihood of your observation for a sequence of coin flips (y_i), here encoded at 0,1 with '1' being 'heads' provided in "Input.txt" via Whiteboard. Compute the **log-likelihood** for $\theta \in (0.3, 0.4, 0.5, 0.6, 0.7)$ and save it as "Exc8Task1c.txt" in the comma-seperated text format using 3 digits after the comma (format '%1.3f'), precisely matching the reference solution.
Call the program "Exc8Task1c.py" and upload via Whiteboard.
d) [programming; print and hand in] Plot the likelihood, as well as the log-likelihood as a **smooth** function of θ (do not interpolate or use any smothing or kernel operations).

Homework 2 (Likelihood - multivariate, (programming counts 2/6 + 2/6))
Consider the Markov Model shown in Fig. 1, with $p_{01} + p_{00} = 1 = p_{11} + p_{10}$ and $0 < p_{(..)} < 1$. This model could e.g. describe the sequence of 'party-events' where an occasional smoker does smoke cigarettes '1' vs. not '0'. The parameter values for this model are not known, but we have a hunch that the person switches from being non-smoker for a while to being smoker and vice versa.
a) [pen & paper] What are the independent parameters of the model θ_j?
b) [pen & paper] Write down the likelihood of the sequence ('00001100111000'), $\mathcal{L}_{(y_i)|\theta}$.
(Tipp: ...think about what the observations y_i are within that sequence, ...)
c) [bonus 2pts] Derive the maximum likelihood estimate θ^* (via pen & paper).
d) [programming; print and hand in] Write a program that computes the likelihood of the observations (y_i) in the sequence provided in the file "Input.txt" via Whiteboard. Plot the **negative log likelihood** of the observations $-\log \mathcal{L}_{(y_i)|\theta}$ as a contour plot. You can choose 'θ_i = np.linspace(0.1,0.9,50)' and plt.contour(X, Y, Z, levels) with levels = np.arange(16, 20, 0.1) for visualization.
e) [programming; submit via whiteboard] Use 'scipy.optimize.minimize' to find the MLE estimate θ^* for the sequence and save it as "Exc8Task2e.txt" in text format using 2 digits after the comma (format '%1.2f') and precisely matching the reference solution where the first number denotes θ_1^*, the second θ_2^* and so on (depending on the number of parameters you identified). Call the program "Exc8Task2e.py" and upload via Whiteboard.
Tipp: Define parameter bounds such that $0 < \theta_i < 1$.

good luck! ...

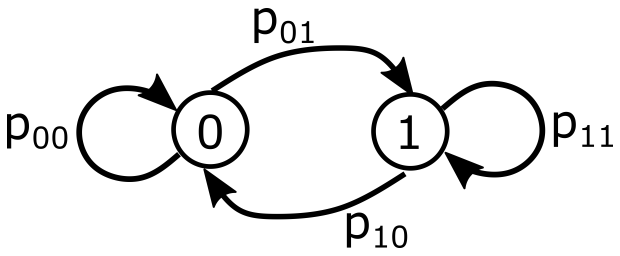

Figure 1: Markov Model