STATISTICAL PRINCIPALS AND COMPUTATIONAL METHODS.
1. SEMINAR – MARKOV CHAINS

Exercise 1. Given the following word model: All words have equal length N. Words are composed of the letters a and b. The first n_a positions are composed of letters a, the remainder is composed of letters b. Words that contain only letters a or only letters b are also valid ($0 \leq n_a \leq N$). See below for a schematic view of such a word.

$$\underbrace{\overbrace{a\ldots a}^{n_a} b\ldots b}_{N}.$$

Informally speaking, the model describes all possible segmentations of a sequence of length N into two parts. Define a Markov model according to which all words described by the above model are equally probable (and all others have zero probability). Define the distribution of states at the first time point $p(y_1)$ and the transition probabilities $p(y_i|y_{i-1})$, $i = 2\ldots N$. Bear in mind that these matrices are not necessarily equal for all i.

Exercise 2. Let the set of states, K, be totally ordered, i.e. $k \in K = \{1, 2, \ldots, m\}$. Let the matrix of transition probabilities $p(y_i = k|y_{i-1} = k')$ be equal for all positions $i = 2\ldots n$, and let it be of the form:

$$p(k \mid k') = \begin{cases} a & \text{for } k = k',\, k' \neq m, \\ b & \text{for } k = k'+1,\, k' \neq m, \\ 1 & \text{for } k = k' = m \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

with $a, b > 0$ and $a + b = 1$. Let the distribution $p(y_1)$ for the start state be 1 for $k = 1$ and 0 otherwise.

a) What is the probability $p(y_i{=}1)$ that the sequence at time i passes through the state $k{=}1$?

b) What is the probability $p(y_i{=}k)$ that the sequence at time i passes through a state $k{\neq}1$?

Exercise 3. Consider the following Markov model with the binary set of states, i.e. $K = \{0,1\}$. Let the matrix of transition probabilities (homogeneous, i.e. equal for all positions) $p(k|k')$ be defined as follows:

$$
\begin{aligned}
p(k{=}0|k'{=}0) &= 1-\alpha \\
p(k{=}1|k'{=}0) &= \alpha \\
p(k{=}0|k'{=}1) &= 1 \\
p(k{=}1|k'{=}1) &= 0
\end{aligned}
$$

Let the distribution for the start state be $p(y_1{=}0) = \beta$ and $p(y_1{=}1) = 1-\beta$.
a) How does the probability $p(y_i{=}0)$ for time point i result from the probability $p(y_{i-1}{=}0)$ for the previous time point $i-1$? Specify the mapping $p(y_{i-1}{=}0) \mapsto p(y_i{=}0)$ explicitly.
b) How does the probability $p(y_i{=}0)$ for time point i result from the probability $p(y_1{=}0) = \beta$ for the first time point? Show that for $i \to \infty$ this probability does not depend on β.
c) What follows for the stationary value $\lim_{i\to\infty} p(y_i{=}0)$?

Exercise 4. In a Markov model for sequences of states y, let one of the states, $k^* \in K$, be distinct. We want to know how often a sequence y generated by this model contains this state on average. Specify an efficient method for calculating this average value.

Hint: Let $n(y)$ denote the number of times the distinct state is contained in the sequence y. The task is to determine the expected value

$$\mathbb{E}_p(n) = \sum_y p(y) \cdot n(y) \; .$$

Furthermore, make use of the fact that the expected value of a sum of random variables is the sum of their expected values! The number of occurrences of k^* in a sequence y can obviously be written as

$$n(y) = \delta_{y_1 k^*} + \delta_{y_2 k^*} + \ldots + \delta_{y_n k^*}$$

(where δ is the Kronecker symbol, i.e. $\delta_{ab} = 1$ if $a = b$, and 0 otherwise).

Exercise 5.
a) A square consists of $n \times n$ many fields. To each field (x, y) we have assigned costs $c(x, y)$. We look at all paths that

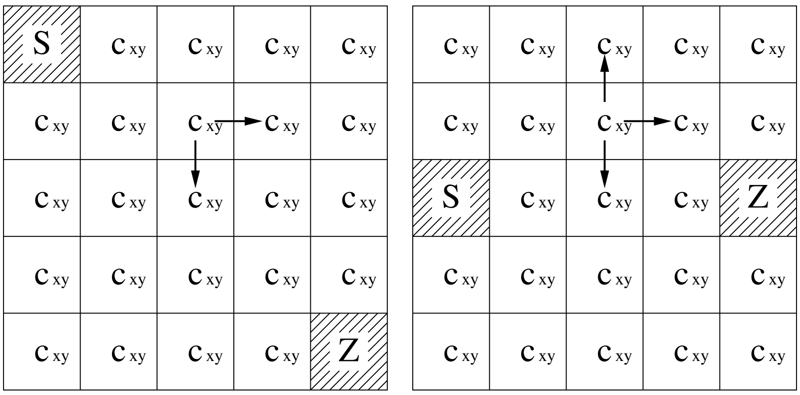
(1) connect the field at the upper left (S) with the field at the lower right (Z), and
(2) where moves can go either towards right or downwards (see Figure (a)).

The total costs of a path is the sum over the costs of all visited fields. Describe (using pseudo-code) an efficient algorithm that finds a path of minimal total cost.

b) Let us slightly modify the task just described:

(1) Now we want to look at all paths that do connect the middle most fields in the first and the last columns, and
(2) moves can now go towards right, towards up, and towards down (see Figure (b)).

Again, describe an efficient algorithm that finds a path of minimal total cost.



(a) (b)

Exercise 6. In the lecture we considered the supervised learning of Markov Chains, where the transition probability matrices $\psi_{i-1,i}$ might be different. Therefore, it was possible to optimize them independently. How the learning should be modified, if these matrices are the same?

Exercise 7. Consider the task of unsupervised parameter learning of an HMM (inhomogeneous, i.e. all functions ψ may be different). Assume that the number of observation sequences in the training set $L = (x^1, x^2 \dots x^l)$ is less or equal the number of states for hidden variables, i.e. $|L| \leq K$. What is the maximal value of the log-Likelihood that can be reached? How to choose the functions ψ in order to reach this value?

Hint: In the case $|L| \leq K$ it is possible to build a HMM model that generates only those sequences that are presented in the training set.