

# Examination Bioinformatics

Linköpings Universitet, IDA, Statistik

---

Course:	732A51 Bioinformatics
Date:	2024/01/09, 8–12
Teacher:	Krzysztof Bartoszek
Provided aids:	The help material is included in the zip file <b>exam_help_material_732A51.zip</b> .
Grades:	A= [18 – 20] points B= [16 – 18) points C= [14 – 16) points D= [12 – 14) points E= [10 – 12) points F= [0 – 10) points
Instructions:	<p>Provide a detailed report that includes plots, conclusions and interpretations. Give motivated answers to the questions. If an answer is not motivated, the points are reduced. Provide all necessary codes in an appendix. If you are asked to do plots, then make sure that they are informative, have correctly labelled axes, informative axes limits and are correctly described. Points may be deducted for poorly done graphs. Name your digital part solution files as: <b>[your exam account id]_[own file description]</b> If you have problems with creating a pdf you may submit your solutions in text files with unambiguous references to graphics and code that are saved in separate files There are <b>THREE</b> assignments (with sub-questions) to solve. Include all code that was used to obtain your answers in your solution files. Make sure it is clear which code section corresponds to which question. Your code should be complete and readable, possible to run by copying directly into a script. Comment directly in the code whenever something needs to be explained or discussed. If you also need to provide some hand-written derivations please number each page according to the pattern: Question number . page in question number i.e. Q1.1, Q1.2, Q1.3,..., Q2.1, Q2.2, ..., Q3.1, ...</p>

---

## Problem 1 (6p)

We consider a HMM with hidden states,  $S_1, S_2$ , on the alphabet  $A = \{1, 2\}$ ,

transition matrix for hidden layer  $P = \begin{bmatrix} 1/3 & 2/3 \\ 0 & 1 \end{bmatrix}$ , initial distribution  $\pi = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ ,

and emission probabilities by the hidden states ( $b_i(j)$  is the probability that a hidden state that equals  $i$  will emit observed state  $j$ ):

$$b_1(1) = 1/3, b_1(2) = 2/3,$$

$$b_2(1) = 0, b_2(2) = 1.$$

What are all possible state sequences for the following observed sequences  $\mathcal{O}$ , and what is the probability of each observed sequence?

(a)  $\mathcal{O} = 1, 1, 2$

(b)  $\mathcal{O} = 1, 2, 2$

(c) Is it possible to generate the observed sequence  $\mathcal{O} = 2, 2, 1$ ?

## Problem 2 (7p)

In the file `gene_expression_measurements.RData` you have a matrix with log intensity measurements from the green and the red channel, from a two sample microarray study. Perform a graphical analysis of the data. Is there a need to do normalization, with a loess curve procedure in this case? Identify (by visual inspection) genes that are outliers. Which ones are up and which ones are down regulated with respect to which sample/channel? Propose a statistical test whether these are differentially regulated. Run this test on these points and check if it is supported that your found genes are differentially regulated.

### Problem 3 (7p)

Consider the graph in Fig. 1 and answer

1. (1p) Write out the adjacency matrix of the graph.
2. (2p) Is the graph bipartite? Justify your answer.
3. (1p) Provide a cycle decomposition of the graph.
4. (1p) What is the diameter of the graph? Assume each edge has unit length.
5. (2p) Which node is the bottleneck/hub in the graph? What is its *betweenness* value? Justify your answer.

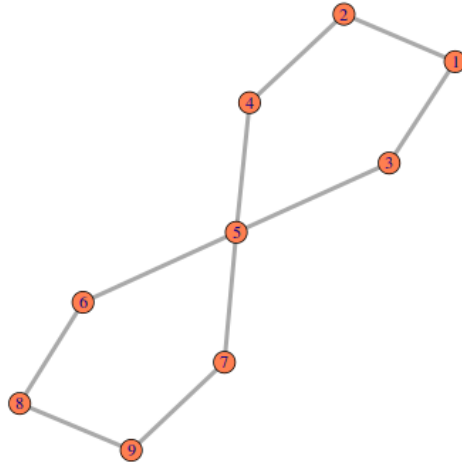


Figure 1: Graph for Problem 3.