

Modeling Cell Fate

Prof. Dr. Volkhard Helms
Maryam Nazarieh, Thorsten Will
Summer Semester 2015

Saarland University
Chair for Computational Biology

Exercise Sheet 2

Due: Jun 02, 2015 10:00 am

Submission

- You are advised to work in groups of two people. If necessary, we will suggest teammates.
- Submit your solutions on paper at the beginning of the lecture in the lecture hall or in Room 3.02, both E2 1. Alternatively you may send an email with a single PDF attachment to maryam.nazarieh@bioinformatik.uni-saarland.de. Late submissions will not be considered.
- If appropriate, include source code listings into the submitted document, we will not merge and layout your source code. If relevant sources are missing on the exercise sheet, they will not be graded.
- Do not forget to mention your names/matriculation numbers.
- Discussion of this exercise will be on Tuesday, Jun 9th at 12:45 in the lecture room (E2 1 007).

Exercise 2.1: Stationary Processes (30 points)

Let Z_t be a sequence of independent normal random variables, each with mean 0 and variance σ^2 , and let a , b , and c be constants. Which, if any, of the following processes are stationary? For each stationary process specify the mean and autocovariance function.

- (a) $X_t = a + bZ_t + cZ_{t-2}$
- (b) $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$
- (c) $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$
- (d) $X_t = a + bZ_0$
- (e) $X_t = Z_0 \cos(ct)$
- (f) $X_t = Z_t Z_{t-1}$

Exercise 2.2: Overrepresentation Analysis (70 points)

- (a) Briefly describe the hypergeometric and fisher exact test. Explain the conditions required to apply each test. Under which conditions, will both tests return the same p-value?
- (b) We have gene expression values for 20,000 genes in two conditions X and Z and found 500 genes differentially expressed. In the entire gene set, 4000 are known to be associated with a particular biological function B. Your computational method detected 100 genes associated with function B in the list of differentially expressed genes **(30 points)**.
 - write a program to calculate the hypergeometric test to assess whether the observation is indeed statistically significant. Report the associated P-value.

Table 1: Hypergeometric Test

Total number of genes	20000
All function B associated genes	4000
Differentially expressed genes	500
Function B associated genes in the list of differentially expressed genes	100

- (c) A sample of people is divided into women and men on the one hand, and those that are healthy and have cancer on the other. We observed that the proportion of patients is higher among the women than among the men (see table below), and we want to test whether the difference of proportions that we observe is significant.

Write a program to calculate Fisher exact test on the following two tables to report whether cancer statistically is related to gender (assuming the significance level is 0.05). Report the P-value, where the null hypothesis that men and women are equally likely to have cancer (**40 points**).

Table 2: Fisher exact Test

	Men	Woman	Row total
Cancer	4	8	12
Non-cancer	8	4	12
Column total	12	12	24

Table 3: Fisher exact Test

	Men	Woman	Row total
Cancer	6	14	20
Non-cancer	14	6	20
Column total	20	20	20

- which table returns a significant P-value?

Have fun!