

Exercises – Lecture 5

Undirected network reconstruction – part 1

Question 1 (*without R*)

Consider a network of four genes A, B, C and D .

- a) The edge set $\{(A, B), (C, D)\}$ specifies the connections among the genes. Draw the corresponding interaction graph for these genes.
- b) In part a) let each edge represent a conditional dependence relationship between the genes, e.g. $A \not\perp B | C, D$. Give all (conditional) independence statements that are implied by this graph.
- c) The edge set $\{(A, B), (B, C), (C, D)\}$ specifies the connections among the genes. Draw the corresponding interaction graph for these genes.
- d) In part c) let each edge represent a conditional dependence relationship between the genes. Give all (conditional) independence statements that are implied by this graph.
- e) The edge set $\{(A, B), (B, C), (C, D), (D, A)\}$ specifies the connections among the genes. Draw the corresponding interaction graph for these genes.
- f) In part f) let each edge represent a conditional dependence relationship between the genes. Give all (conditional) independence statements that are implied by this graph.

Question 2 (*without R*)

Let the bivariate random variable $\mathbf{Y} = (Y_1, Y_2)^\top$ be distributed as:

$$\mathbf{Y} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right).$$

- a) Give the correlation matrix of the random variable \mathbf{Y}
- b) Describe in words the differences between the distribution of \mathbf{Y} and the following bivariate normal distribution:

$$\mathbf{Z} \sim \mathcal{N}\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\right).$$

Also draw a plot of how data according to both distributions are distributed.

Question 3 (*without R*)

Let $X \sim \mathcal{N}(0, 1)$ and $E \sim \mathcal{N}(0, 1)$ be independent, and let $Y = X + \frac{1}{2}E$. Calculate the correlation between X and Y . *Hint:* First calculate the variance of Y , then the covariance of X and Y .

Question 4 (*without R*)

Consider a two-gene pathway. The expression levels of gene A, denoted Y_A , are distributed as

$\mathcal{N}(2, 1)$. Gene A has a repressing effect on gene B. Their relation between their expression levels is given by the linear regression model: $Y_B = 3 - \frac{3}{4}Y_A + \varepsilon_B$, with $\varepsilon_B \sim \mathcal{N}(0, \frac{1}{4})$. Furthermore, assume Y_A and ε_B are independent.

- a) Give the conditional and unconditional mean of Y_B .
- b) Give the conditional and unconditional variance of Y_B .
- c) Calculate the covariance between Y_A and Y_B .
- d) Calculate the correlation between Y_A and Y_B .
- e) Using the answers to the above questions write down the bivariate normal distribution of $(Y_A, Y_B)^\top$.

Question 5 (*without R*)

Consider a two-gene pathway that processes a signal S . The two genes react differently to the signal as can be witnessed from the following system of regression equations describing their expression levels:

$$\begin{aligned} Y_a &= -S + \gamma\varepsilon_a, \\ Y_b &= S + \gamma\varepsilon_b, \end{aligned}$$

where γ a positive constant and S, ε_a and ε_b independent and all distributed as $\mathcal{N}(0, 1)$.

- a) Calculate the correlation between the (expression levels of the) two genes.
- b) What happens to the correlation calculated in part a) when $\gamma \rightarrow \infty$? How can this be understood in terms of the influence of the signal S ?

Question 6 (*with R*)

The expression levels of the genes of a two-gene pathway have been measured in five individuals. The data are provided in the table below:

Individual	Y_1	Y_2
1	1.748	0.638
2	1.658	1.238
3	1.962	1.344
4	2.496	1.837
5	-3.677	-1.493

- a) Assume the data above are bivariate normal distributed. Estimate the parameters of this distribution from these data.
- b) Still assume bivariate normality and (in addition) the independence of Y_a and Y_b . Re-estimate the parameters of the bivariate normal distribution incorporating the independence assumption.
- c) From the answer of part a) also calculate the correlation between the expression levels of the two genes. Confirm a nonzero correlation by means of regression analysis.