

Exercises – Lecture 7

Undirected network reconstruction - part 3

For the exercises the following formula's may be useful.

The inverse of a 2×2 matrix \mathbf{A} is:

$$\mathbf{A}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = [\det(\mathbf{A})]^{-1} \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix}$$

with $\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$.

The inverse of a 3×3 matrix \mathbf{A} is:

$$\begin{aligned} \mathbf{A}^{-1} &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}^{-1} \\ &= [\det(\mathbf{A})]^{-1} \begin{pmatrix} a_{33}a_{22} - a_{32}a_{23} & -(a_{33}a_{12} - a_{32}a_{13}) & a_{23}a_{12} - a_{22}a_{13} \\ -(a_{33}a_{21} - a_{31}a_{23}) & a_{33}a_{11} - a_{31}a_{13} & -(a_{23}a_{11} - a_{21}a_{13}) \\ a_{32}a_{21} - a_{31}a_{22} & -(a_{32}a_{11} - a_{31}a_{12}) & a_{22}a_{11} - a_{21}a_{12} \end{pmatrix} \end{aligned}$$

with $\det(\mathbf{A}) = a_{11}(a_{33}a_{22} - a_{32}a_{23}) - a_{21}(a_{33}a_{12} - a_{32}a_{13}) + a_{31}(a_{23}a_{12} - a_{22}a_{13})$.

Question 1 (without R)

Consider a five-gene pathway. The inverse covariance matrix of the expression levels of the five genes is given by:

$$\Sigma^{-1} = \begin{pmatrix} 1 & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 2 & \frac{3}{4} & 1 & \frac{3}{4} \\ 0 & \frac{3}{4} & 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 & 1 & \frac{1}{10} \\ 0 & \frac{3}{4} & 0 & \frac{1}{10} & 1 \end{pmatrix}$$

Draw the graph depicting the conditional (in)dependencies between the five genes of the pathway. Use as a criterion to draw an edge only absence or presence of the edge, and do not worry about the 'strength' of a conditional dependency.

Question 2 (without R)

Consider a 3-gene pathway. Expression levels of the three genes in the pathway follow a multivariate

normal distribution:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 & 2 & -2 \\ 2 & 5 & -5 \\ -2 & -5 & 8 \end{pmatrix}\right)$$

Calculate the partial correlations between each pair of genes in this pathway. Using the partial correlations draw a graph reflecting the independence relationships between the genes.

Question 3 (without R)

The topology of a 3-gene pathway is given by the graph of (in)dependence relations with edge set $\mathcal{E} = \{(1, 2), (2, 3)\}$. The corresponding (conditional) distributions of the genes' expression levels are: $Y_1 | Y_2, Y_3 \sim \mathcal{N}(1, 4)$, $Y_2 | Y_1, Y_3 \sim \mathcal{N}(\frac{1}{2}Y_1 - \frac{7}{2}, 4)$, and $Y_3 | Y_1, Y_2 \sim \mathcal{N}(1 - Y_2, 3)$.

- a) Give the joint (multivariate) distribution of (Y_1, Y_2, Y_3) . *Hint:* first determine the joint (multivariate) distribution of (Y_1, Y_2) .
- b) Calculate the partial correlations between each pair of genes.
- c) Write the relationships between the genes as a system of three regression equations. In each equation the expression levels of one gene are regressed on those of the other genes, e.g.: $Y_1 = \beta_{1,2}Y_2 + \beta_{1,3} + \varepsilon_1$. Use the multivariate distribution obtained above to specify the values of the regression coefficients in these equations. *Hint:* hereto exploit the relationship between the regression coefficients and the partial correlations (as given in the lecture notes).

Question 4

Consider a three-gene pathway. Expression data of this pathway is available for 100 samples. Below are the result of three regression analyses. In each regression analysis (the expression levels of) one gene is regressed on (the expression levels of) the other genes.

Regression of gene 1 on the other genes:

```

Coefficients :
              Estimate Std.Error t-value Pr(> |t|)
gene 2      -0.24071    0.03872   -6.217 1.24e-08 ***
gene 3       0.72330    0.03904   18.526 < 2e-16 ***
  
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 0.6742 on 98 degrees of freedom
Multiple R-squared: 0.7779, Adjusted R-squared: 0.7734
F-statistic: 171.6 on 2 and 98 DF, p-value: < 2.2e-16
  
```

Regression of gene 2 on the other genes:

```

Coefficients :
              Estimate Std.Error t-value Pr(> |t|)
gene 1      -1.175     0.189   -6.217 1.24e-08 ***
gene 3       1.093     0.146    7.483 3.17e-11 ***
  
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.49 on 997 degrees of freedom
 Multiple R-squared: 0.3636, Adjusted R-squared: 0.3506
 F-statistic: 27.99 on 2 and 98 DF, p-value: 2.416e-10

Regression of gene 3 on the other genes:

Coefficients :					
	Estimate	Std.Error	t - value	Pr(> t)	
gene 1	1.07547	0.05805	18.526	< 2e - 16	***
gene 2	0.33280	0.04448	7.483	3.17e - 11	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8221 on 98 degrees of freedom
 Multiple R-squared: 0.8029, Adjusted R-squared: 0.7989
 F-statistic: 199.6 on 2 and 98 DF, p-value: < 2.2e-16

- (without R) Draw the dependency graph between the three genes on the basis of the regression analyses. In this use the 'traditional' cut-off: $p < 0.05$.
- (with R) Give the partial correlation matrix.
- (with R) Test which of these partial correlations differs significantly from zero.
- (with R) Give the covariance matrix from which these data were generated.

Question 5* (not necessarily with R)

The topology of the conditional independence graph of a 3-gene pathway is given by the edge set: $\{(1, 2), (1, 3), (2, 3)\}$. The corresponding (conditional) distributions of the genes' expression levels are: $Y_1 | Y_2 \sim \mathcal{N}(-Y_2, \frac{1}{2})$, $Y_2 | Y_1 \sim \mathcal{N}(-\frac{1}{2}Y_1, \frac{1}{4})$, and $Y_3 | Y_1, Y_2 \sim \mathcal{N}(Y_1 + Y_2, 1)$. Finally, the partial errors, e.g. the errors in Y_1 conditional on Y_2 and Y_3 , are independent.

- Give the covariance matrix of (Y_1, Y_2, Y_3) . *Hint*: first determine the covariance matrix of (Y_1, Y_2) .
- Verify that none of the partial correlations are zero.

Question 6* (without R)

Consider the following partitioned covariance matrix:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

The corresponding inverse covariance matrix $\Omega = \Sigma^{-1}$ can be partitioned in the same way. Show that, if $\Sigma_{12} = \mathbf{0} = \Sigma_{21}^T$, the corresponding blocks of Ω are also zero. In other words, show that (in this particular case) marginal independence implies conditional independence.

Question 7

Consider a four-gene pathway. Expression data of this pathway is available for 1000 samples. Below are the result of four regression analyses. In each regression analysis (the expression levels of)

one gene is regressed on (the expression levels of) the other genes.

Regression of gene 1 on the other genes:

Coefficients :				
	Estimate	Std.Error	t - value	Pr(> t)
gene 2	-0.48543	0.03427	-14.166	< 2e - 16 ***
gene 3	0.03006	0.03735	0.805	0.421
gene 4	-0.03074	0.04397	-0.699	0.485

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7074 on 997 degrees of freedom
Multiple R-squared: 0.3641, Adjusted R-squared: 0.3622
F-statistic: 190.3 on 3 and 997 DF, p-value: < 2.2e-16

Regression of gene 2 on the other genes:

Coefficients :				
	Estimate	Std.Error	t - value	Pr(> t)
gene 1	-0.34517	0.02437	-14.17	< 2e - 16 ***
gene 3	-0.65077	0.02383	-27.31	< 2e - 16 ***
gene 4	-0.36067	0.03528	-10.22	< 2e - 16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5965 on 997 degrees of freedom
Multiple R-squared: 0.6915, Adjusted R-squared: 0.6906
F-statistic: 745.1 on 3 and 997 DF, p-value: < 2.2e-16

Regression of gene 3 on the other genes:

Coefficients :				
	Estimate	Std.Error	t - value	Pr(> t)
gene 1	0.02160	0.02684	0.805	0.421
gene 2	-0.65772	0.02408	-27.314	< 2e - 16 ***
gene 4	-0.02214	0.03727	-0.594	0.553

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5997 on 997 degrees of freedom
Multiple R-squared: 0.5853, Adjusted R-squared: 0.584
F-statistic: 469 on 3 and 997 DF, p-value: < 2.2e-16

Regression of gene 4 on the other genes:

Coefficients :				
	Estimate	Std.Error	t - value	Pr(> t)
gene 1	-0.01594	0.02280	-0.699	0.485
gene 2	-0.26307	0.02573	-10.223	< 2e - 16 ***
gene 3	-0.01598	0.02690	-0.594	0.553

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5095 on 997 degrees of freedom

Multiple R-squared: 0.2105, Adjusted R-squared: 0.2081

F-statistic: 88.62 on 3 and 997 DF, p-value: < 2.2e-16

- a) (*without R*) Based on the regression analyses above draw the dependency graph between the four genes. In this use the 'traditional' cut-off: $p < 0.05$.
- b) (*with R*) Give the partial correlation matrix.
- c) (*with R*) Test which of these partial correlations differs significantly from zero.
- d) (*with R*) Give the covariance matrix from which these data were generated.