# Answers – Lecture 5
# Undirected network reconstruction – part 1

**Question 1**
Graph omitted in answer, should be obvious.

*Question 1b)*
$A \perp\!\!\!\perp C$; $A \perp\!\!\!\perp D$; $B \perp\!\!\!\perp C$; $B \perp\!\!\!\perp D$;
$A, B \perp\!\!\!\perp C, D$; $A, B \perp\!\!\!\perp C$; $A, B \perp\!\!\!\perp D$; $A \perp\!\!\!\perp C, D$; $B \perp\!\!\!\perp C, D$;
$A \perp\!\!\!\perp C \,|\, B$; $A \perp\!\!\!\perp C \,|\, D$; $A \perp\!\!\!\perp D \,|\, B$; $A \perp\!\!\!\perp D \,|\, C$; $B \perp\!\!\!\perp C \,|\, A$; $B \perp\!\!\!\perp C \,|\, D$; $B \perp\!\!\!\perp D \,|\, A$; $B \perp\!\!\!\perp D \,|\, C$;
$A \perp\!\!\!\perp C \,|\, B, D$; $A \perp\!\!\!\perp D \,|\, B, C$; $B \perp\!\!\!\perp C \,|\, A, D$; $B \perp\!\!\!\perp D \,|\, A, C$.

*Question 1d)*
$A \perp\!\!\!\perp C \,|\, B$; $A \perp\!\!\!\perp D \,|\, B$; $A \perp\!\!\!\perp C \,|\, B, D$; $A \perp\!\!\!\perp D \,|\, C$; $A \perp\!\!\!\perp D \,|\, B, C$;
$B \perp\!\!\!\perp D \,|\, C$; $B \perp\!\!\!\perp D \,|\, A, C$.

*Question 1f)*
$A \perp\!\!\!\perp C \,|\, B, D$; $B \perp\!\!\!\perp D \,|\, A, C$.

**Question 2**
*Question 2a)*
The correlation matrix is:

$$\begin{aligned} \mathbf{R} &= \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}^{-1} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}. \end{aligned}$$

Alternatively, use the fact that the diagonal of the correlation matrix contains only ones. And the off-diagonal elements are given by $\rho(Y_1, Y_2) = \text{Cov}(Y_1, Y_2)/\{[\text{Var}(Y_1)]^{1/2}[\text{Var}(Y_2)]^{1/2}\} = 1/(\sqrt{2}\sqrt{2}) = 1/2$.

*Question 2b)*
The mean and correlation have changed. Visually, this results in a shift of the location of the data and also a rotation of 'sigar-shaped' relation between the variates of $\mathbf{Y}$.

**Question 3**
The variance of $Y$: $\text{Var}(Y) = \text{Var}(X + \frac{1}{2}E) = \text{Var}(X) + \text{Var}(\frac{1}{2}E) = \text{Var}(X) + \frac{1}{4}\text{Var}(E) =$

$1 + \frac{1}{4}$, where we have used the independence between $X$ and $E$. The covariance of $X$ and $Y$: $\mathrm{Cov}(X,Y) = \mathrm{Cov}(X, X + \frac{1}{2}E) = \mathrm{Cov}(X,X) + \mathrm{Cov}(X, \frac{1}{2}E) = \mathrm{Var}(X) + \frac{1}{2}\mathrm{Cov}(X,E) = 1 + 0 = 1$, where again we have used the independence between $X$ and $Y$. The correlation is now: $\mathrm{Cov}(X,Y)/[\mathrm{Var}(X)\mathrm{Var}(Y)]^{1/2} = \sqrt{\frac{4}{5}}$.

**Question 4**

*Question 4a)*

For the conditional mean $Y_A$ is temporarily fixed at some value, say $Y_A = y_a$. Hence, $Y_A$ is temporarily non-random, equalling some constant. As a consequence, the only randomness in $Y_B$, which equals $3 - \frac{3}{4}Y_A + \varepsilon_B$ given $Y_A$, is in $\varepsilon_B$. Hence, the expectation of $Y_B$ given $Y_A$ is only with respect to $\varepsilon_B$. Thus: $E_{\varepsilon_B}(Y_B \,|\, Y_A = y_A) = E_{\varepsilon_B}(3 - \frac{3}{4}y_A + \varepsilon_B) = E_{\varepsilon_B}(3) - E_{\varepsilon_B}(\frac{3}{4}y_A) + E_{\varepsilon_B}(\varepsilon_B) = 3 - \frac{3}{4}y_A$.

For the unconditional mean of $Y_B$ let $Y_A$ be random again. The mean of $Y_B$ is now to be calculated taking into account both the uncertainty of $\varepsilon_B$ and $Y_A$. Then $E_{Y_A}[E_{\varepsilon_B}(Y_B \,|\, Y_A)] = E_{Y_A}[3 - \frac{3}{4}Y_A] = 3 - \frac{3}{4}E_{Y_A}[Y_A] = 3 - \frac{3}{4} \cdot 2 = 3/2$.

Alternatively:

$$\begin{aligned}
E(Y_B) &= E_{Y_A}(E_{\varepsilon_B}(Y_B \,|\, Y_a = y_A)) \\
&= \int_{-\infty}^{\infty} E_{\varepsilon_B}(Y_B \,|\, Y_a = y_A) f_{Y_A}(y_A) dy_A \\
&= \int_{-\infty}^{\infty} (3 - \frac{3}{4}y_A) f_{Y_A}(y_A) dy_A \\
&= 3\int_{-\infty}^{\infty} f_{Y_A}(y_A) dy_A - \frac{3}{4}\int_{-\infty}^{\infty} y_A f_{Y_A}(y_A) dy_A \\
&= 3 - \frac{3}{4}E(Y_A) \\
&= \frac{3}{2}.
\end{aligned}$$

*Question 4b)*

Conditional variance: $\mathrm{Var}(Y_B \,|\, Y_A) = \mathrm{Var}(3 - \frac{3}{4}Y_A + \varepsilon_B \,|\, Y_A = y_a) = \mathrm{Var}(3 \,|\, Y_A = y_a) + \mathrm{Var}(-\frac{3}{4}Y_A \,|\, Y_A = y_a) + \mathrm{Var}(\varepsilon_B \,|\, Y_A = y_a) = \mathrm{Var}(3) + \mathrm{Var}(-\frac{3}{4}y_A) + \mathrm{Var}(\varepsilon_B) = 0 + 0 + \frac{1}{4} = \frac{1}{4}$.

Alternatively, note that because of the independence of $Y_A$ and $\varepsilon_B$ we have $f_{\varepsilon_B \,|\, Y_A} = f_{\varepsilon_B, Y_A}/f_{Y_A} = f_{\varepsilon_B}\, f_{Y_A}/f_{Y_A} = f_{\varepsilon_B}$. Then:

$$\begin{aligned}
\mathrm{Var}(Y_B \,|\, Y_A = y_A) &= \int_{-\infty}^{\infty} [3 - \frac{3}{4}y_A + \varepsilon_B - E_{\varepsilon_B}(Y_B \,|\, Y_a = y_A)]^2 \, f_{\varepsilon_B \,|\, Y_A = y_a}(\varepsilon_B, y_A)\, d\varepsilon_B \\
&= \int_{-\infty}^{\infty} \varepsilon_B^2 \, f_{\varepsilon_B}(\varepsilon_B)\, d\varepsilon_B \\
&= \mathrm{Var}(\varepsilon_B) \\
&= \frac{1}{4}.
\end{aligned}$$

Unconditional variance:

$$\begin{aligned}
\mathrm{Var}(Y_B) &= \mathrm{Var}(3 - \frac{3}{4}Y_A + \varepsilon_B) \\
&= \mathrm{Var}(3) + \mathrm{Var}(-\frac{3}{4}Y_A) + \mathrm{Var}(\varepsilon_B) \\
&= \frac{9}{16}\mathrm{Var}(Y_A) + \frac{1}{4} \\
&= \frac{9}{16} + \frac{1}{4} \\
&= \frac{13}{16},
\end{aligned}$$

in which we have used the independence between $Y_A$ and $\varepsilon_B$.

Alternatively:

$$\begin{aligned}
\mathrm{Var}(Y_B) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [3 - \frac{3}{4}y_A + \varepsilon_B - E(Y_B))]^2 \, f_{\varepsilon_B, Y_A}(\varepsilon_B, y_A)\, d\varepsilon_B\, dY_A \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [3 - \frac{3}{4}y_A + \varepsilon_B - 3 + \frac{3}{4}E(Y_A)]^2 \, f_{Y_A}(y_A)\, f_{\varepsilon_B}(\varepsilon_B)\, d\varepsilon_B\, dY_A \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [-(\frac{3}{4}y_A - \frac{3}{4}E(Y_A)) + \varepsilon_B]^2 \, f_{Y_A}(y_A)\, f_{\varepsilon_B}(\varepsilon_B)\, d\varepsilon_B\, dY_A \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\frac{9}{16}(y_A - E(Y_A))^2 + \varepsilon_B^2 - \frac{6}{4}(y_A - E(Y_A))\varepsilon_B] \, f_{Y_A}(y_A)\, f_{\varepsilon_B}(\varepsilon_B)\, d\varepsilon_B\, dY_A \\
&= \frac{9}{16}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y_A - E(Y_A))^2 + \varepsilon_B^2 \, f_{Y_A}(y_A)\, f_{\varepsilon_B}(\varepsilon_B)\, d\varepsilon_B\, dY_A \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varepsilon_B^2 \, f_{Y_A}(y_A)\, f_{\varepsilon_B}(\varepsilon_B)\, d\varepsilon_B\, dY_A \\
&\quad - \frac{6}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y_A - E(Y_A))\varepsilon_B \, f_{Y_A}(y_A)\, f_{\varepsilon_B}(\varepsilon_B)\, d\varepsilon_B\, dY_A \\
&= \frac{9}{16}\mathrm{Var}(Y_A) + \mathrm{Var}(\varepsilon_B) \\
&= \frac{9}{16} + \frac{1}{4} \\
&= \frac{13}{16},
\end{aligned}$$

as before.

*Question 4c)*
Covariance:

$$\begin{aligned}
\mathrm{Cov}(Y_A, Y_B) &= \mathrm{Cov}(Y_A, 3 - \frac{3}{4}Y_A + \varepsilon_B) \\
&= \mathrm{Cov}(Y_A, 3) - \mathrm{Cov}(Y_A, \frac{3}{4}Y_A) + \mathrm{Cov}(Y_A, \varepsilon_B) \\
&= -\frac{3}{4}\mathrm{Var}(Y_A) \\
&= -\frac{3}{4},
\end{aligned}$$

in which we have used the independence between $Y_A$ and $\varepsilon_B$.

*Question 4d)*
Correlation:

$$\begin{aligned} \text{Cor}(Y_A, Y_B) &= \text{Cov}(Y_A, Y_B)/[\text{Var}(Y_A)\text{Var}(Y_B)]^{1/2} \\ &= -\frac{3}{4}/\sqrt{\frac{13}{16}} \end{aligned}$$

*Question 4e)*
The bivariate distribution is given by:

$$\begin{pmatrix} Y_A \\ Y_B \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 2 \\ \frac{3}{2} \end{pmatrix}, \begin{pmatrix} 1 & -\frac{3}{4} \\ -\frac{3}{4} & \frac{13}{16} \end{pmatrix} \right).$$

**Question 5**
*Question 5a)*
The covariance between $Y_a$ ad $Y_b$ is:

$$\begin{aligned} \text{Cov}(Y_a, Y_b) &= \text{Cov}(-S + \gamma\,\varepsilon_a, S + \gamma\,\varepsilon_b) \\ &= \text{Cov}(-S + \gamma\,\varepsilon_a, S + \gamma\,\varepsilon_b) \\ &= \text{Cov}(-S, S) + \text{Cov}(-S, \gamma\,\varepsilon_b) + \text{Cov}(\gamma\,\varepsilon_a, S) + \text{Cov}(\gamma\,\varepsilon_a, \gamma\,\varepsilon_a) \\ &= -\text{Var}(S, S) \;=\; -1. \end{aligned}$$

Similarly,

$$\begin{aligned} \text{Var}(Y_a) &= \text{Cov}(-S, -S) + \ldots + \text{Cov}(\gamma\,\varepsilon_a, \gamma\,\varepsilon_a) \\ &= 1 + \gamma^2 \\ &= \text{Var}(Y_b). \end{aligned}$$

Together this yields: $\text{Cov}(Y_a, Y_b) = -1/(1 + \gamma^2)$.

*Question 5b)*
The correlation tends to zero as $\gamma \to \infty$. As $\gamma \to \infty$ the larger the contribution of $\gamma\,\varepsilon$: the larger the noise. As a consequence, the signal drowns in the noise, and its influence on $Y_a$ and $Y_b$ fades. The signal is shared by $Y_a$ and $Y_b$ and as its influence vanishes so will the correlation.

**Question 6**
*Question 6a)*
Assume the object `Y` contains the data with rows and columns corresponding to individuals and genes, respectively. Then, the mean parameters are estimated as:

```
> mean(Y[,1])
> mean(Y[,2])
```

and the covariance matrix by:

```
> cov(Y)
```

*Question 6b)*
In this bivariate case the incorporation of the independence assumption amounts to setting *rho* equal to zero. This can be verified by univariate estimation of the parameters. For the mean the `R`-code is given above, for the variance:

```
> var(Y[,1])
> var(Y[,2])
```

*Question 6c)*
To calculate the correlation matrix:

```
> cov(Y) .
```

This clearly deviates from zero. Regression analysis confirms a non-zero relation between $Y_1$ and $Y_2$:

```
> summary(Y[,1] ∼ Y[,2]) .
```