Graphical Models: Worksheet 4 MT 2023

Questions will not be marked, however solutions will be provided.

A: Warm Up

A1. Causal Models

Let $p^*(y \mid x) = \int p(y \mid x, z)p(z)\, dz$.

(a) Draw a causal graph with vertices X, Y, Z for which $p^*(y \mid x) = p(y \mid do(x))$.
This is the graph in which $Z \to X \to Y$ and $Z \to Y$.

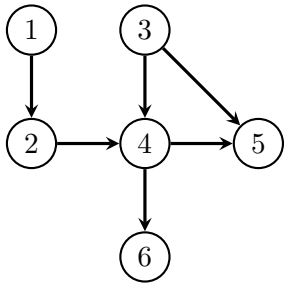
(b) Show that p^* is a valid conditional distribution for Y given $X = x$. (In other words, show that it is non-negative and integrates to 1 for each fixed x).
It is clearly non-negative since the integrand is non-negative. Integrating with respect to y and swapping the order of the integration shows that it integrates to 1.

(c) Show that $p^*(y \mid x) = p(y \mid x)$ if *either* $X \perp\!\!\!\perp Z$ or $Y \perp\!\!\!\perp Z \mid X$.

If $X \perp\!\!\!\perp Z$ then $p(z) = p(z \mid x)$, so $\int p(y \mid x, z)p(z)\, dz = \int p(y \mid x, z)p(z \mid x)\, dz = \int p(y, z \mid x)\, dz = p(y \mid x)$. If $Y \perp\!\!\!\perp Z \mid X$ then $p(y \mid x, z)$ doesn't depend upon z, so $p^*(y \mid x) = p(y \mid x) \int p(z)\, dz = p(y \mid x)$.

A2. d-Separation

Consider the DAG below.



Which of the following d-separation statements are true?

(i) $2 \perp_d 3 \mid 4$; *False since the path $2 \to 4 \leftarrow 3$ is open conditional on 4. However $2 \perp_d 3 \mid \emptyset$ is true.*

(ii) $2 \perp_d 5 \mid 4$; *False since the path $2 \to 4 \leftarrow 3 \to 5$ is open conditional on 4. However $2 \perp_d 5 \mid 3, 4$ is true.*

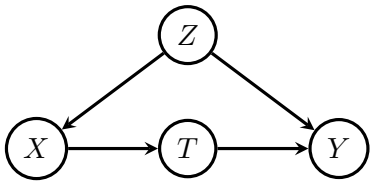
(iii) $1 \perp_d 5 \mid 3, 4$; *True since all paths out of 5 start with a non-collider (3 or 4), and these are both in the conditioning set.*

(iv) $5 \perp_d 6 \mid 4$. *True, since 4 is a non-collider on both paths from 5 to 6.*

For those that are not true, identify an open path, and also a separating set for which the statement *is* true.

B: Core Questions

B1. Front-Door Adjustment



Assume that p is Markov with respect to the graph $\mathcal{G}$ shown above, and that $(\mathcal{G}, p)$ represents a causal model.

(a) Show that

$$p(y \mid do(x)) = \sum_t p(t \mid x) \sum_z p(y \mid z, t) p(z).$$

We have

$$\begin{aligned} p(y \mid do(x)) &= \sum_{z,t} \frac{p(z, x, t, y)}{p(x \mid z)} \\ &= \sum_{z,t} p(y \mid z, t) p(t \mid x) p(z) \\ &= \sum_t p(t \mid x) \sum_z p(y \mid z, t) p(z) \end{aligned}$$

as required.

(b) Show further that

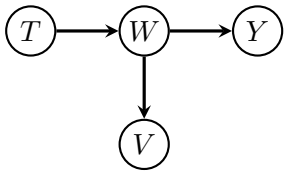
$$\sum_z p(y \mid z, t) p(z) = \sum_x p(y \mid x, t) p(x),$$

and hence deduce a formula for $p(y \mid do(x))$ that does not involve Z. *[Hint: write $p(z) = \sum_x p(z \mid x) \cdot p(x)$, and use the conditional independences from the graph.] Note that*

$$\begin{aligned} \sum_z p(y \mid z, t) p(z) &= \sum_z p(y \mid z, t) \sum_x p(x, z) \\ &= \sum_z \sum_x p(y \mid z, t) p(z \mid x) p(x) \\ &= \sum_{z,x} p(y \mid x, z, t) p(z \mid x, t) p(x) \\ &= \sum_{z,x} p(y, z \mid x, t) p(x) \\ &= \sum_x p(y \mid x, t) p(x). \end{aligned}$$

It follows that $p(y \mid do(x)) = \sum_t p(t \mid x) \sum_{x'} p(y \mid x', t) p(x')$ (which is the same as $\sum_t p(t \mid x) \cdot p(y \mid do(t))$).

B2. Regression Adjustment



For this question, assume that we are dealing with a Gaussian causal model, so that the causal effect is just $\beta_{ty \cdot C}$, where C is any valid adjustment set.

(a) Consider the graph above. Show that:

$$\beta_{ty \cdot v} = \beta_{tw} \cdot \beta_{wy} \cdot \frac{\sigma_{vv \cdot w}}{\sigma_{vv \cdot t}} = \beta_{tw} \cdot \beta_{wy} \cdot \frac{d_{vv}}{d_{vv} + \beta_{wv}^2 d_{ww}},$$

where d_{vv} is the variance of the error term for the structural equation generating X_v.

First note that since $\sigma_{tt \cdot v}\sigma_{vv} = \sigma_{tt}\sigma_{vv} - \sigma_{tv}^2$, we have $\sigma_{tt \cdot v}\sigma_{vv} = \sigma_{vv \cdot t}\sigma_{tt}$. Also, by the trek rule we have that $\sigma_{vy} = d_{ww}\beta_{wv}\beta_{wy} + d_{tt}\beta_{tw}^2\beta_{wv}\beta_{wy}$ and also $\sigma_{ww} = d_{ww} + d_{tt}\beta_{tw}^2$; hence $\sigma_{vy} = \sigma_{ww}\beta_{wv}\beta_{wy}$. Then

$$\begin{aligned}
\beta_{ty \cdot v} &= \frac{\sigma_{ty \cdot v}}{\sigma_{tt \cdot v}} = \frac{\sigma_{ty} - \sigma_{tv}\sigma_{vy}/\sigma_{vv}}{\sigma_{vv \cdot t}\sigma_{tt}/\sigma_{vv}} \\
&= \frac{\sigma_{tt}\beta_{tw}\beta_{wy} - \sigma_{tt}\beta_{tw}\beta_{wv}\sigma_{vy}/\sigma_{vv}}{\sigma_{vv \cdot t}\sigma_{tt}/\sigma_{vv}} \\
&= \beta_{tw}\frac{\beta_{wy}\sigma_{vv} - \beta_{wv}\sigma_{ww}\beta_{wy}\beta_{wv}}{\sigma_{vv \cdot t}} \\
&= \beta_{tw}\beta_{wy}\frac{\sigma_{vv} - \beta_{wv}^2\sigma_{ww}}{\sigma_{vv \cdot t}} \\
&= \beta_{tw}\beta_{wy}\frac{\sigma_{vv \cdot w}}{\sigma_{vv \cdot t}},
\end{aligned}$$

where the last equality follows from the fact that $\beta_{wv} = \sigma_{wv}/\sigma_{ww}$. Then note that $\sigma_{vv \cdot w}$ is just d_{vv} by the definition of the least squares equation, and $\sigma_{vv \cdot t} = \sigma_{vv} - \beta_{tv}^2\sigma_{tt} = d_{vv} + d_{ww}\beta_{wv}^2$ using the trek rule and the fact that T is exogenous.

(b) Deduce that, if $v \in C$ then adjustment on X_C does not give a consistent estimate of the causal effect $T \to Y$.

Note that the true causal effect is just $\beta_{tw}\beta_{wy}$, and the fraction is clearly < 1 if V is correlated with W (and if the distribution is not degenerate), so clearly V is not a valid adjustment set. Clearly we cannot include W in a valid adjustment set, so the only other candidates are $\{V\}$ (which we have ruled out) and $\emptyset$. Hence, in this particular graph V is not a member of any valid adjustment set.

(c) Use this to deduce that, for *any* causal DAG $\mathcal{G}$, no member of $\text{forb}_{\mathcal{G}}(T \to Y)$ can belong to an adjustment set that gives a consistent estimate for most distributions.

There are two types of member of $\text{forb}_{\mathcal{G}}(T \to Y)$, things that are in $\text{cn}_{\mathcal{G}}(T \to Y)$ and their strict descendants. Clearly adjusting for something on the causal path will not be consistent, since the effect could be entirely mediated by that variable and our estimate would be 0. For any descendant (say V), we can pick a

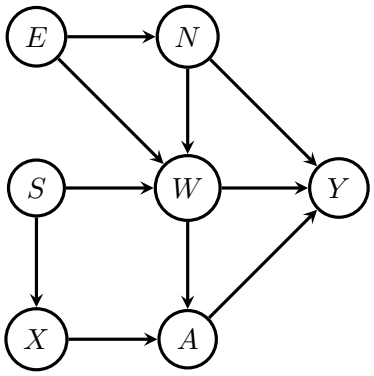
distribution in the model so that the entire effect is mediated through the ancestor in $\text{cn}_{\mathcal{G}}(T \to Y)$*, and have a single non-trivial path down to* V*. (If there is a vertex earlier on the path in the candidate set, then use that instead.) All other variables can be chosen to be independent. Then the previous analysis shows that the estimate of the effect is (in general) biased if we adjust with a set containing* V*.*

B3. Adjustment Sets

A cardiologist is interested in the mechanisms which cause a myocardial infarction (heart attack, Y). She believes that it is directly affected by the patient's diet (N), their weight (W), and the build up of fat in their arteries (A). The patient's weight is also an effect of their diet, and a cause of fat in their arteries. Weight and diet are each affected by the patient's socio-economic status (E), and weight is also a function of their sex (S). Suppose also that the doctor has access to a new drug, X, which she assigns at random conditional on the patient's sex, and whose only effect is to reduce the arterial fat build up.

(a) Draw a directed acyclic graph over the seven variables, that minimally represents the causal story described.

The graph should be:



(b) List all the valid adjustment sets for the causal effect of W on Y.

This would be:

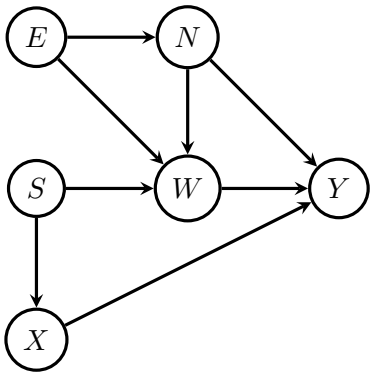
$$\{N, S\} \qquad \{N, X\} \qquad \{N, X, S\}$$
$$\{N, S, E\} \qquad \{N, X, E\} \qquad \{N, X, S, E\}.$$

(c) Suppose that we assume a linear model for each of the variables conditional upon their parents. Which of the valid adjustment sets is likely to lead to the estimate of the causal effect that has the lowest variance, and why?

By Henckel's Theorem (Theorem 8.34), we know that the lowest variance belongs to the set

$$\begin{aligned} O_{\mathcal{G}}(W \to Y) &= \text{pa}_{\mathcal{G}}(\text{cn}_{\mathcal{G}}(W \to Y)) \setminus (\{W\} \cup \text{cn}_{\mathcal{G}}(W \to Y)) \\ &= \{X, W, A, N\} \setminus \{W, A, Y\} \\ &= \{X, N\}. \end{aligned}$$

(d) Compute the forbidden projection for (W, Y), and hence verify your answer.
The graph should be:

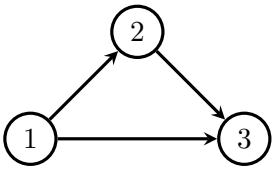


and hence $\mathrm{pa}_{\widetilde{\mathcal{G}}}(Y) \setminus \{W\} = \{X, N\}$ *as required.*

C: Optional

C1. Mediation

Let $\mathcal{G}$ be the graph shown below, and suppose that $(X_1, X_2, X_3)^T \sim N(0, \Sigma)$ is causal with respect to the graph below.



Let $\beta_{ij \cdot A} := \sigma_{ij \cdot A}/\sigma_{ii \cdot A}$ be the coefficient of the variable X_i when performing a linear regression of X_j on X_i, X_A. Note that this quantity does **not** depend upon any causal structure.

(a) Write $\beta_{13 \cdot 2}$ in terms of entries of Σ.
Using the definition of the Schur complement, we have

$$\begin{aligned}\beta_{13 \cdot 2} &:= (\sigma_{13} - \sigma_{12}\sigma_{23}/\sigma_{22})/(\sigma_{11} - \sigma_{12}^2/\sigma_{22}) \\ &= (\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23})/(\sigma_{22}\sigma_{11} - \sigma_{12}^2).\end{aligned}$$

(b) Using the trek rule, show directly that $b_{21} = \beta_{12}$, and $b_{31} = \beta_{13 \cdot 2}$.
The trek rule gives $\sigma_{11} = d_{11}$, *and*

$$\begin{aligned}\sigma_{12} &= d_{11}b_{21}, \qquad \sigma_{13} = d_{11}b_{21}b_{32} + d_{11}b_{31} \\ \sigma_{23} &= d_{11}b_{21}b_{31} + d_{11}b_{21}^2 b_{32} + d_{22}b_{32}.\end{aligned}$$

Hence $\beta_{12} = \sigma_{12}/\sigma_{11} = b_{21}$ *as required. The bottom of the ratio above for* $\beta_{13 \cdot 2}$ *is*

$$\sigma_{22}\sigma_{11} - \sigma_{12}^2 = (d_{22} + b_{21}^2 d_{11})d_{11} - (d_{11}b_{21})^2 = d_{22}d_{11}.$$

The top is

$$\begin{aligned}
&\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23} \\
&= (d_{22} + b_{21}^2 d_{11})(d_{11}b_{21}b_{32} + d_{11}b_{31}) - d_{11}b_{21}(d_{11}b_{21}b_{31} + d_{11}b_{21}^2 b_{32} + d_{22}b_{32}) \\
&= d_{22}(d_{11}b_{21}b_{32} + d_{11}b_{31}) - d_{11}b_{21}d_{22}b_{32}) \\
&= d_{22}d_{11}b_{31}.
\end{aligned}$$

This gives the result.

(c) Argue that if i is a parent of j then $b_{ji} = \beta_{ij \cdot A}$, where $A = \mathrm{pa}_{\mathcal{G}}(j) \setminus \{i\}$.
[Hint: see Sheet 3 qB3.]
Using part (c) of the question suggested, we see that X_j can be written as a linear function of its parents (with coefficients b_{ji} and an independent error term); hence the regression coefficients are just as described.

(d) Show that $\sigma_{13} = \sigma_{13 \cdot 2} + \sigma_{12}\sigma_{23}/\sigma_{22}$ and hence (or otherwise) deduce that

$$\beta_{13} = \beta_{13 \cdot 2} + \beta_{12}\beta_{23 \cdot 1}.$$

[Hint: let $r_3 := X_3 - \beta_{13 \cdot 2}X_1 - \beta_{23 \cdot 1}X_2$, and recall that $\beta_{13 \cdot 2}, \beta_{12 \cdot 1}$ are defined so that r_3 is uncorrelated with both X_1 and X_2.]
Using the bilinearity of covariance (i.e. it is linear in each of its arguments),

$$\begin{aligned}
\mathrm{Cov}(X_1, X_3) &= \mathrm{Cov}(X_1, X_3 - \beta_{23 \cdot 1}X_2) + \mathrm{Cov}(X_1, \beta_{23 \cdot 1}X_2) \\
&= \mathrm{Cov}(X_1, \beta_{13 \cdot 2}X_1) + \beta_{23 \cdot 1}\,\mathrm{Cov}(X_1, X_2) \\
\sigma_{13} &= \sigma_{11}\beta_{13 \cdot 2} + \sigma_{12}\beta_{23 \cdot 1},
\end{aligned}$$

where $\mathrm{Cov}(X_1, X_3 - \beta_{23 \cdot 1}X_2) = \mathrm{Cov}(X_1, \beta_{13 \cdot 2}X_1)$ follows from the hint that $\mathrm{Cov}(X_1, X_3 - \beta_{23 \cdot 1}X_2 - \beta_{13 \cdot 2}X_1) = 0$.
Note that the same result can also be deduced from the trek rule in the graph shown above.
In a system such as $\mathcal{G}$, the first term of this formula is sometimes called the *direct effect* of X_1 on X_3, and the second term the *indirect effect* via X_2.

(e) Can you separate out causal effects in a more general way? For example, consider partitioning into paths of length $l \geq 1$.
We have

$$\beta_{1k} = \sum_{l \geq 1} \sum_{i_0 < i_1 < i_2 < \cdots < i_l} \prod_{s=1}^{l} b_{i_{s-1}, i_s},$$

This is (almost) just the trek rule applied to a graph in which the first variable has no parents, and hence all treks are just directed paths from 1 to k.

C2. Causal Effects

The *average causal effect* on Y of changing $Z = z$ to $Z = z'$ is defined as

$$\mathrm{ACE}_{Z \to Y}(z', z) := \mathbb{E}[Y \mid do(Z = z')] - \mathbb{E}[Y \mid do(Z = z)].$$

(a) Show that if $(\mathcal{G}, p)$ is causal and $p(x_V)$ is a multivariate Gaussian distribution, then

$$\mathrm{ACE}_{i\to j}(x_i', x_i) := \beta_{i\to j}(x_i' - x_i)$$

where $\beta_{i\to j}$ is the regression coefficient of X_i when regressing X_j on X_i, X_B for any valid adjustment set B.

From the beginning of Section 8.6 and the fact that B is a valid adjustment set, we know that $\mathbb{E}[X_j \mid do(x_i)]$ can be obtained by averaging a regression model for X_j given X_i, X_B over $p(x_B)$. It follows that $\mathbb{E}[X_j \mid do(x_i)] = \alpha + \beta_{i\to j} x_i$ for some constants $\alpha, \beta_{i\to j}$; it follows from the same derivation that $\beta_{i\to j}$ is the coefficient of X_i in that regression. Hence, $\mathbb{E}[X_j \mid do(x_i')] - \mathbb{E}[X_j \mid do(x_i)] = \beta_{i\to j}(x_i' - x_i)$.

(b) Show further that

$$\beta_{i\to j} := \sum_{\pi \in \mathcal{D}_{ij}} \prod_{k\to l \in \pi} b_{lk},$$

where $\mathcal{D}_{ij}$ is the set of directed paths from i to j. *[Hint: consider the quantity $\mathrm{Cov}(X_j, X_i - \sum_{c\in \mathrm{pa}_\mathcal{G}(i)} b_{ic} X_c)$ and use the trek rule.]*

Let $C = \mathrm{pa}_\mathcal{G}(i)$; this is a valid adjustment set so we can write

$$\beta_{i\to j} = \frac{\mathrm{Cov}(X_j, X_i - \sum_{c\in \mathrm{pa}_\mathcal{G}(i)} b_{ic} X_C)}{\mathrm{Var}(X_i - \sum_{c\in \mathrm{pa}_\mathcal{G}(i)} b_{ic} X_C)}.$$

[This is like regressing X_i on its parents and then regressing X_j on the residual.] Now,

$$\mathrm{Cov}(X_j, X_i - \sum_{c\in \mathrm{pa}_\mathcal{G}(i)} b_{ic} X_C) = \mathrm{Cov}(X_j, X_i) - \sum_{c\in \mathrm{pa}_\mathcal{G}(i)} b_{ic} \mathrm{Cov}(X_j, X_c).$$

Now, by the trek rule, the first term includes all treks from j to i, while the sum removes precisely treks from i to j that begin with an edge $i \leftarrow c$. This leaves only one-sided treks with source i, i.e. directed paths from i to j. This is identical to the expression given, except for a factor of d_{ii}.

However,

$$\mathrm{Var}(X_i - \sum_{c\in \mathrm{pa}_\mathcal{G}(i)} b_{ic} X_C) = \mathrm{Var}(\varepsilon_i) = d_{ii}$$

(in the usual notation), so this gives the result.

C3. Forbidden Projection

Prove Theorem 8.40, stating that if $\widetilde{\mathcal{G}}$ is the forbidden projection of $\mathcal{G}$ with respect to (T, Y), then

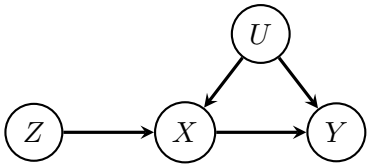
$$O_\mathcal{G}(T \to Y) = \mathrm{pa}_{\widetilde{\mathcal{G}}}(Y) \setminus \{T\}.$$

First we show that $O_\mathcal{G}(T \to Y) \subseteq \mathrm{pa}_{\widetilde{\mathcal{G}}}(Y) \setminus \{T\}$. Suppose that $v \in O_\mathcal{G}(T \to Y) = \mathrm{pa}_\mathcal{G}(\mathrm{cn}_\mathcal{G}(T \to Y)) \setminus (\mathrm{cn}_\mathcal{G}(T \to Y) \cup \{T\})$, so it is a parent of a causal node. Since every $\mathrm{cn}_\mathcal{G}(T \to Y)$ is in $\mathrm{forb}_\mathcal{G}(T \to Y)$, this means that there is a directed path from v to y such that all intermediate nodes are projected out in $\widetilde{\mathcal{G}}$.

For the converse, note that $\mathrm{pa}_{\widetilde{\mathcal{G}}}(Y)$ *consists of nodes that were previously (strict) ancestors of* Y*, but such that the nodes on the directed path to* Y *have been removed in* $\widetilde{\mathcal{G}}$*. This is precisely the definition of an element of* $O_{\mathcal{G}}(T \to Y)$*, because the ancestors of* Y *that are removed are precisely the elements of* $\mathrm{cn}_{\mathcal{G}}(T \to Y) \setminus \{Y\}$*, and the immediate strict parents of this set* $\mathrm{cn}_{\mathcal{G}}(T \to Y)$ *are precisely* $O_{\mathcal{G}}(T \to Y) \cup \{T\}$*.*

C4. Instrumental Variables

Consider the four Gaussian variable causal system shown.



(a) Show that, if $\mathrm{Cov}(Z, X) \neq 0$, we have $\beta_{X \to Y} = \mathrm{Cov}(Z, Y)/\mathrm{Cov}(Z, X)$.

Using the trek rule, we have $\mathrm{Cov}(Z, Y) = d_{zz} b_{xz} b_{yx}$*, and* $\mathrm{Cov}(Z, X) = d_{zz} b_{xz}$*. Then, provided* $\mathrm{Cov}(Z, X) \neq 0$*, (so in particular* $b_{xz} \neq 0$*) this gives the result.*

(b) Explain the utility of this result if U is unobserved.

The formula provided only involves the other three variables, so we can obtain an estimate of the causal effect of X *on* Y *even in the presence of unobserved confounders.*

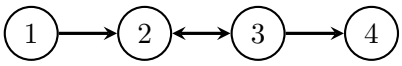
Note that this result relies strongly on there being no direct effect of Z *on* Y*, nor any correlation between* U *and* Z*.*

In the literature you may see this described as 'two-stage least squares' (2SLS) because we perform two ordinary linear regressions (Y *on* Z *and* X *on* Z*) to get our estimate.*

C5. Correlated Errors

In our formulation of Gaussian DAGs we found that the error terms were independent (see question B3 on Sheet 3), and hence the matrix $D = \mathrm{Cov}(\boldsymbol{\varepsilon})$ is diagonal. One possible extension to this model is to allow for *correlated errors*, i.e. so that D is an arbitrary covariance matrix.

We can represent this graphically by including a *bidirected* edge $(i \leftrightarrow j)$ whenever $d_{ij} = d_{ji} \neq 0$.



(a) Consider the graph shown. Evaluate $(I - B)^{-1}$.

Following the usual derivation we have $(I - B)^{-1} = I + B + B^2 + B^3$*. In this case* $B^2 = 0$*, so we just get* $I + B$*, i.e.*

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ b_{21} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & b_{43} & 1 \end{pmatrix}$$

(b) Hence derive Σ in terms of b_{21}, b_{43} and non-zero entries of D.

We have

$$\begin{aligned}
\Sigma &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ b_{21} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & b_{43} & 1 \end{pmatrix} \begin{pmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & d_{23} & 0 \\ 0 & d_{23} & d_{33} & 0 \\ 0 & 0 & 0 & d_{44} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ b_{21} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & b_{43} & 1 \end{pmatrix}^T \\
&= \begin{pmatrix} d_{11} & 0 & 0 & 0 \\ d_{11}b_{21} & d_{22} & d_{23} & 0 \\ 0 & d_{23} & d_{33} & 0 \\ 0 & d_{23}b_{43} & d_{33}b_{43} & d_{44} \end{pmatrix} \begin{pmatrix} 1 & b_{21} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & b_{43} \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} d_{11} & d_{11}b_{21} & 0 & 0 \\ d_{11}b_{21} & d_{22} + d_{11}b_{21}^2 & d_{23} & d_{23}b_{43} \\ 0 & d_{23} & d_{33} & d_{33}b_{43} \\ 0 & d_{23}b_{43} & d_{33}b_{43} & d_{44} + d_{33}b_{43}^2 \end{pmatrix}
\end{aligned}$$

(c) Derive an analogue of the trek rule that applies to graphs with correlated errors of this form.

As you might guess from the derivation above, we now need to include as a trek the possibility of the source *being a bidirected edge. For example, in the graph in the question, the entry for σ_{22} consists of the usual treks two from 2 to itself. However, for $\sigma_{24} = d_{23}b_{43}$ this looks like a trek with source $2 \leftrightarrow 3$, left hand side consisting of the trivial path 2, and right hand side consisting of $3 \to 4$.*