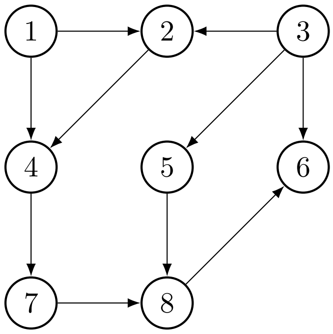
MODERN STATISTICAL METHODS Part III
Example Sheet 4 (of 4) RDS/Michaelmas 2025

Questions 4 and 10 will be marked.

1.



Write down:

(a) all sets $S \subseteq [8] \setminus \{1,3\}$ of variables that d-separate 1 and 3;

(b) all sets $S \subseteq [8] \setminus \{1,4,6\}$ of variables that d-separate $\{1,4\}$ and 6.

2. The *skeleton* of a DAG $\mathcal{G}$ is the undirected graph obtained by ignoring the directions of the edges in the DAG. A *v-structure* in a DAG is a triple j, l, k of nodes with $j \to l \leftarrow k$ where j and k are not adjacent. Explain why if for a DAG $\mathcal{G}$, we have $j \perp\!\!\!\perp_{\mathcal{G}} k \,|\, S$ and we have edges $j - l$ and $k - l$ in the skeleton with $l \notin S$, then j, l, k must be a v-structure.

3. In this question we will outline an algorithm to compute the graphical Lasso.

(a) Let
$$Q(\Omega) = -\log\det(\Omega) + \operatorname{tr}(S\Omega) + \lambda\|\Omega\|_1$$
be the graphical Lasso objective with $\hat{\Omega} = \underset{\Omega \succ 0}{\operatorname{argmin}}\, Q(\Omega)$ assumed unique. Consider the following version of the graphical Lasso objective:
$$\min_{\Omega,\Theta \succ 0} \{-\log\det(\Omega) + \operatorname{tr}(S\Omega) + \lambda\|\Theta\|_1\}$$
subject to $\Omega = \Theta$. By introducing the Lagrangian for this objective, show that
$$p + \max_{U: S+U \succ 0,\, \|U\|_\infty \le \lambda} \log\det(S+U) \le Q(\hat{\Omega}).$$
Here $\|U\|_\infty = \max_{j,k} |U_{jk}|$, $A \succ 0$ means matrix A is positive definite and p is the number of columns in the underlying data matrix X. [*Hint: Use the fact that* $\nabla \log\det(\Omega) = \Omega^{-1}$ *and write the additional term in the Lagrangian as* $\operatorname{tr}(U(\Omega - \Theta))$.]

(b) Suppose that U^* is the unique maximiser of the LHS of the previous display. Show that $\hat{\Omega} = (S + U^*)^{-1}$.

(c) Now consider
$$\hat{\Sigma} = \underset{W: W \succ 0,\, \|W - S\|_\infty \le \lambda}{\operatorname{argmin}} -\log\det(W). \tag{1}$$

By using the formula for the determinant in terms of Schur complements, show that $(\hat\Sigma_{jj}, \hat\Sigma_{-j,j}) = (\alpha^*, \beta^*)$, where (α^*, β^*) solve the following optimisation problem over (α, β):

$$\begin{aligned} &\text{minimise} && -\alpha + \beta^T \hat\Sigma_{-j,-j}^{-1} \beta, \\ &\text{such that} && \|\beta - S_{-j,j}\|_\infty \le \lambda,\ |\alpha - S_{jj}| \le \lambda. \end{aligned}$$

Conclude that $\alpha^* = S_{jj} + \lambda$.

(d) Show that $\beta^* = \hat\Sigma_{-j,-j}\theta^*$, where $\theta^* \in \mathbb{R}^{p-1}$ is the unique minimiser of the Lasso-type objective function

$$\theta \mapsto \frac{1}{2}\theta^\top \hat\Sigma_{-j,-j}\theta - \theta^\top S_{-j,j} + \lambda\|\theta\|_1$$

over $\mathbb{R}^{p-1}$. (Thus, fixing $\hat\Sigma_{-j,-j}$, we may find optimal values for the remaining parts of $\hat\Sigma$, which suggests that by cycling through j we can perform a form of blockwise coordinate descent.)

4. Assume the setup of Theorem 40.

(a) Show that

$$\begin{aligned} &\left|\hat\varepsilon_i^2 \hat\xi_i^2 - \varepsilon_i^2 \xi_i^2\right| \\ &\quad\le \left\{\left(\hat f(Z_i) - f(Z_i)\right)^2 + 2\left|\varepsilon_i\left(\hat f(Z_i) - f(Z_i)\right)\right|\right\}\left\{\left(\hat g(Z_i) - g(Z_i)\right)^2 + 2\left|\xi_i\left(\hat g(Z_i) - g(Z_i)\right)\right|\right\} \\ &\qquad + \varepsilon_i^2\left\{\left(\hat g(Z_i) - g(Z_i)\right)^2 + 2\left|\xi_i\left(\hat g(Z_i) - g(Z_i)\right)\right|\right\} \\ &\qquad + \xi_i^2\left\{\left(\hat f(Z_i) - f(Z_i)\right)^2 + 2\left|\varepsilon_i\left(\hat f(Z_i) - f(Z_i)\right)\right|\right\} \\ &\quad =: \mathrm{I}_i + \mathrm{II}_i + \mathrm{III}_i. \end{aligned}$$

(b) Show that

$$\begin{aligned} \mathrm{I}_i \le\ & 3\left(\hat f(Z_i) - f(Z_i)\right)^2\left(\hat g(Z_i) - g(Z_i)\right)^2 + \varepsilon_i^2\left(\hat g(Z_i) - g(Z_i)\right)^2 + \xi_i^2\left(\hat f(Z_i) - f(Z_i)\right)^2 \\ &+ 4\left|\varepsilon_i\xi_i\left(\hat f(Z_i) - f(Z_i)\right)\left(\hat g(Z_i) - g(Z_i)\right)\right|. \end{aligned}$$

(c) Show that

$$A^2 := \frac{1}{n}\sum_{i=1}^n \left(\hat f(Z_i) - f(Z_i)\right)^2\left(\hat g(Z_i) - g(Z_i)\right)^2 \xrightarrow{p} 0.$$

(d) Show that

$$B^2 := \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2(\hat g(Z_i) - g(Z_i))^2 \xrightarrow{p} 0.$$

(e) Complete the proof of Theorem 40 by showing that $\tau_D \xrightarrow{p} \sqrt{\mathrm{Var}(\varepsilon\xi)}$. [*Hint: Appeal to symmetry where possible, to avoid unnecessary calculations.*]

5. Prove Theorem 41. [*Hint: Define*

$$\hat\tau_1^{(k)} := \frac{1}{n}\sum_{i \in I_k} \frac{(Y_i - \hat\mu_1^{(k)}(x_i))T_i}{\hat\pi^{(k)}(X_i)} + \hat\mu_1^{(k)}(X_i),$$

and $\hat\tau_1^{*,(k)}$ *as above, but with* $\hat\pi^{(k)}$ *and* $\hat\mu_1^{(k)}$ *replaced by their targets* π *and* μ_1*; and define* $\hat\tau_0^{(k)}$ *and* $\hat\tau_0^{*,(k)}$ *similarly. Argue that it suffices to show* $\sqrt{n}(\hat\tau_j^{(k)} - \tau_j^{*,(k)})$ *for* $j = 0, 1$, *and use the decomposition from lectures.*]

In the following questions, suppose there are m null hypotheses being tested, $H_1, \dots, H_m$, and let $p_1, \dots, p_m$ be the associated p-values, and let $p_{(1)} \le \cdots \le p_{(m)}$ be the ordered p values (so (i) is the index of the ith smallest p-value). Furthermore, unless otherwise stated, let I_0 be the set of true null hypotheses.

6. Show that the definition of Holm's procedure as the closed testing procedure with the local tests as the Bonferroni test is equivalent to the step-down procedure definition.

7. The Benjamini–Hochberg procedure allows us to control the FDR when the p-values of true null hypotheses are independent of each other, and independent of the false null hypotheses (though see also Qu 10). The following variant of the method, known as the *Benjamini–Yekutieli* procedure allows us to control the FDR under arbitrary dependence of the p-values, and works as follows. Define

$$\gamma_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}.$$

Let $\hat{k} = \max\{i : p_{(i)} \le \alpha i/(m\gamma_m)\}$ and reject $H_{(1)}, \dots, H_{(\hat{k})}$. First show that the FDR of this procedure satisfies

$$\mathrm{FDR} = \sum_{i \in I_0} \mathbb{E}\left(\frac{1}{R}\mathbb{1}_{\{p_i \le \alpha R/(m\gamma_m)\}}\mathbb{1}_{\{R>0\}}\right).$$

Now go on to prove that $\mathrm{FDR} \le \alpha m_0/m \le \alpha$. *Hint: Verify that that for any $r \in \mathbb{N}$ we have*

$$\frac{1}{r} = \sum_{j=1}^{\infty} \frac{\mathbb{1}_{\{j \ge r\}}}{j(j+1)},$$

and use this to replace $1/R$.

8. Consider the closed testing procedure applied to m hypotheses $H_1, \dots, H_m$. Let $\mathcal{R}$ be the collection of all $I \subseteq \{1, \dots, m\}$ for which for all $J \supseteq I$, the local test $\phi_J = 1$. Now suppose that (perhaps after having looked at the results of the ϕ_I), we decide we want to reject a set of hypotheses indexed by $B \subseteq \{1, \dots, m\}$. Let

$$t_\alpha(B) = \max\{|I| : I \subseteq B,\ I \notin \mathcal{R}\}.$$

Show that $\{0, 1, \dots, t_\alpha(B)\}$ gives a $1-\alpha$ confidence set for the number of false rejections in B. That is, show that

$$\mathbb{P}(|B \cap I_0| > t_\alpha(B)) \le \alpha,$$

and that this is true no matter how B is chosen. *Hint: Argue by working on the event* $\{\phi_{I_0} = 0\}$.

9. Let $\mathcal{I}$ be a non-empty subset of the set of subsets of $[m]$, and consider a family of intersection hypotheses $\{H_I : I \in \mathcal{I}\}$ that is hierarchical in the sense that for any $I,\ J \in \mathcal{I}$, we either have $I \cap J = \emptyset$ or $I \subseteq J$ or $J \subseteq I$. Suppose that for each H_I with $I \in \mathcal{I}$ we have a p-value p_I. Define the adjusted p-value of H_I to be

$$p_I^{\mathrm{adj}} := \max_{J \in \mathcal{I} : J \supseteq I} \frac{m}{|J|} p_J.$$

Meinshausen's procedure rejects all hypotheses H_I for which $p_I^{\mathrm{adj}} \le \alpha$. Show that this procedure controls the FWER.

10. (a) A set $D \in [0,1]^d$ is *increasing* if whenever $x \in D$ and $y \in [0,1]^d$ are such that $x_j \leq y_j$ for all $j \in [d]$, we have $y \in D$. Fix $i \in [m]$. Explain why the set of p-values in $[0,1]^{m-1}$ resulting in at most $r-1$ rejections from the modified Benjamini–Hochberg procedure introduced in the proof of Theorem 44 is an increasing set.

(b) Assume now that for each $i \in I_0$ and any increasing set $D \in [0,1]^{m-1}$, we have that $x \mapsto \mathbb{P}(p_{-i} \in D \,|\, p_i \leq x)$ is increasing on $[0,1]$. Prove that the Benjamini–Hochberg procedure controls the FDR. [*Hint: Aim to use the given assumption to obtain a telescoping sum.*]