MODERN STATISTICAL METHODS Part III
Example Sheet 4 (of 4) Michaelmas 2019

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

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

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

   (c) Now consider
   $$\hat{\Sigma} = \underset{W: W \succ 0,\, \|W - S\|_\infty \leq \lambda}{\mathrm{argmin}} -\log\det(W). \tag{1}$$
   Let $\hat{\Sigma}_{-j,j}$ be a block of $\hat{\Sigma}$ containing all but the jth row of $\hat{\Sigma}$ and only the jth column. Use the formula for the determinant in terms of Schur complements to show that $(\hat{\Sigma}_{jj}, \hat{\Sigma}_{-j,j}) = (\alpha^*, \beta^*)$, where (α^*, β^*) solve the following optimisation problem over (α, β):
   $$\begin{aligned} &\text{minimise} && -\alpha + \beta^T \hat{\Sigma}^{-1}_{-j,-j}\beta, \\ &\text{such that} && \|\beta - S_{-j,j}\|_\infty \leq \lambda,\ |\alpha - S_{jj}| \leq \lambda. \end{aligned}$$
   Conclude that $\alpha^* = S_{jj} + \lambda$. ($\beta^*$ can be found by standard quadratic programming techniques, or by converting the optimisation to a standard Lasso optimisation problem; thus we can perform block coordinate descent on the optimisation problem in (1), updating a row and corresponding column of W at each iteration.)

2. Consider the matrix
   $$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1.01 \end{bmatrix}$$
   and its perturbation
   $$\hat{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1.01 \end{bmatrix} + \begin{bmatrix} 0 & 0.01 \\ 0.01 & 0 \end{bmatrix}.$$
   Show that the eigenvalues are stable to the perturbation, but the top eigenvector is not.

3. Prove the following inequality for two rank-r matrices S, U with orthonormal columns,
   $$\min_{O \in \mathbb{R}^{d\times d} \text{ orthogonal}} \|S - UO\|_F \leq \sqrt{2}\sin\Theta(S,U).$$
   Hint: Use $O = VW^T$ where $W\Sigma V^T$ is a singular value decomposition of S^TU.

4. The SCoTTLASS estimator for Sparse PCA is obtained by solving

$$\underset{v \in S^{d-1}}{\text{maximise}}\ v^T \hat{\Sigma} v \quad \text{subject to } \|v\|_1 \leq \lambda.$$

Show that this is equivalent to the optimisation problem

$$\underset{\Theta \in \mathcal{S}_+^{d \times d}}{\text{maximise}}\ \text{Tr}(\Theta\hat{\Sigma}) \quad \text{subject to Tr}(\Theta) = 1, \sum_{i,j} |\Theta_{i,j}| \leq \lambda^2, \text{rank}(\Theta) = 1,$$

where $\mathcal{S}_+^{d \times d}$ is the cone of positive semidefinite matrices. Which of the constraints in this problem are convex? Dropping the rank constraint yields the problem

$$\underset{\Theta \in \mathcal{S}_+^{d \times d}}{\text{maximise}}\ \text{Tr}(\Theta\hat{\Sigma}) \quad \text{subject to Tr}(\Theta) = 1, \sum_{i,j} |\Theta_{i,j}| \leq R^2.$$

What happens when the maximum is achieved by a rank 1 matrix?

5. Consider the spiked covariance model $\Sigma = \theta v v^T + I_d$ with a sparse spike $v \in S^{d-1}$, $\|v\|_0 = k$. Suppose that $k \leq d/2$ and $k \leq k'$ and define the Sparse PCA estimator

$$\hat{v} = \underset{v \in S^{d-1}, \|v\|_0 \leq k'}{\text{argmax}}\ v^T \hat{\Sigma} v.$$

In this problem you shall prove a non-asymptotic error bound for this estimator (Proposition 34 in the notes).

(a) Let $s \subseteq \{1, \ldots, d\}$ be the random subset of entries where either v or $\hat{v}$ is non-zero. Prove that

$$\sin\Theta(v, \hat{v})^2 \leq \frac{1}{\theta}\langle \hat{\Sigma}_s - \Sigma_s, \hat{v}_s\hat{v}_s^T - v_s v_s^T\rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product, and subscripting by s indicates rows (and columns, for a matrix) are subsetted by s.

(b) Use Problem 3 and the matrix inequality $\langle A, B\rangle \leq \|A\|_{op}\|B\|_1$, where $\|B\|_1$ is the Schatten-1 norm, to show that

$$\min_{\epsilon \in \{-1,1\}} \|v - \epsilon\hat{v}\| \leq \frac{2\sqrt{2}}{\theta}\|\hat{\Sigma}_s - \Sigma_s\|_{op}.$$

(c) Finally, apply a union bound over all subsets $w \subseteq \{1, \ldots, d\}$ of size $k + k'$ in conjunction with Theorem 29 in the notes, to show that for some constant c,

$$\mathbb{P}\left(\min_{\epsilon \in \{-1,1\}} \|v - \epsilon\hat{v}\| \geq c\frac{1+\theta}{\theta}\left[\sqrt{\eta_n} \vee \eta_n\right]\right) \leq e^{-\delta} \qquad \text{for all } \delta > 0,$$

with

$$\eta_n = \frac{(k+k')\log(de^2/(k+k')) + \delta}{n}.$$

6. Show that if all null hypotheses are true, then the FDR is equivalent to the FWER.

7. 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 the step-down procedure definition. *Hint: It may help to first show that with the definition as a closed testing procedure, we reject $H_{(i)}$ when*

$$\min_{s \in \{1,\ldots,m\}} \mathbb{1}_{\{\min_{j \in J_s} p_j \leq \alpha/s\}} = 1,$$

where

$$J_s \in \operatorname*{argmin}_{J:J\supseteq\{(i)\},\,|J|=s} \mathbb{1}_{\{\min_{j\in J} p_j \le \alpha/s\}}.$$

8. 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. 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 $H_{(1)}, \ldots, H_{(\hat{k})}$. First show that the FDR of this procedure satisfies
$$\text{FDR} = \sum_{i\in I_0} \mathbb{E}\left(\frac{1}{R}\mathbb{1}_{\{p_i \le \alpha R/(m\gamma_m)\}}\mathbb{1}_{\{R>0\}}\right).$$
Now go on to prove that $\text{FDR} \le \alpha m_0/m \le \alpha$. *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 $1/R$.

9. Consider the closed testing procedure applied to m hypotheses $H_1, \ldots, H_m$. Let $\mathcal{R}$ be the collection of all $I \subseteq \{1, \ldots, m\}$ 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 $B \subseteq \{1, \ldots, m\}$. Let
$$t_\alpha(B) = \max\{|I| : I \subseteq B,\ I \notin \mathcal{R}\}.$$
Show that $\{0, 1, \ldots, t_\alpha(B)\}$ 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* $\{\phi_{I_0} = 0\}$.

In the following questions, let all quantities be as defined in Section 4.3 of the lecture notes concerning the debiased Lasso.

10. Show that
$$(\hat{\Theta}\hat{\Sigma}\hat{\Theta}^T)_{jj} = \frac{1}{n}\|X_j - X_{-j}\hat{\gamma}^{(j)}\|_2^2/\hat{\tau}j^4.$$

11. Show that
$$\frac{1}{n}X_j^T(X_j - X_{-j}\hat{\gamma}^{(j)}) = \frac{1}{n}\|X_j - X_{-j}\hat{\gamma}^{(j)}\|_2^2 + \lambda_j\|\hat{\gamma}^{(j)}\|_1.$$

12. Prove that $\mathbb{P}(\Lambda_n) \to 1$, where the sequence of events Λ_n is defined in the proof of Theorem 40.