Combinatorial and Algebraic Statistics

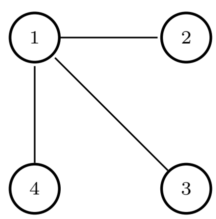
Problem Set 1 version 1.1

Due Date: March 19

1. (a) Let $X \sim \mathcal{N}(0,1)$ be a random variable distributed according to the standard normal distribution. Show that X has expectation $\mathbb{E}[X] = 0$ and variance $\mathrm{Var}[X] = 1$.

   (b) Let $Y = (X_1, \ldots, X_m)$ be a random vector for which the expectations $\mathbb{E}[X_1], \ldots, \mathbb{E}[X_m]$ and $\mathbb{E}[X_i X_j]$ are all finite, for all $i, j \in [m]$. Show that the covariance matrix $\mathrm{Cov}[Y]$ is positive semidefinite.

2. Consider the simplicial complex $\Gamma = [12][13][14]$.



   (a) Show that Γ is decomposable.

   (b) Determine an explicit Markov basis for the hierarchical log-linear model associated with Γ and $r_1 = r_3 = 1,\ r_2 = r_4 = 2$.

3. Consider the m-way independence model $\mathcal{M}_\Gamma$ where $\Gamma = [1][2]\cdots[m]$.

   (a) Find a necessary and sufficient condition for an element of the set $D(\Gamma_1, \Gamma_2)$ to be nonzero for any reducible decomposition (Γ_1, S, Γ_2) of Γ.

   (b) Let (Γ_1, S, Γ_2) and $(\Gamma_1', S', \Gamma_2')$ be two distinct reducible decompositions of Γ. Show that $D(\Gamma_1, \Gamma_2) \cap D(\Gamma_1', \Gamma_2')$ is equal to

$$\left\{ \begin{bmatrix} i & j & k \\ i' & j & k' \end{bmatrix} - \begin{bmatrix} i & j & k' \\ i' & j & k \end{bmatrix} : i, i' \in \mathcal{R}_{[m]\setminus A}, j \in \mathcal{R}_B, k, k' \in \mathcal{R}_C \right\}$$

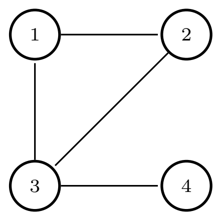
   Where $A = \mathcal{G}(\Gamma_2) \cup \mathcal{G}(\Gamma_2') \cup S \cup S'$, $B = (\mathcal{G}(\Gamma_2) \triangle \mathcal{G}(\Gamma_2')) \cup S \cup S'$, and $C = (\mathcal{G}(\Gamma_2) \cap \mathcal{G}(\Gamma_2')) \setminus (S \cup S')$.

4. We consider Gaussian models associated with undirected graphs on 4 nodes. Let us assume that we have 3 samples:

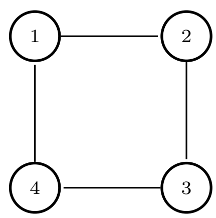
$$Y_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad Y_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad Y_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

The resulting sample covariance matrix is $S = \frac{1}{3}(Y_1 Y_1^T + Y_2 Y_2^T + Y_3 Y_3^T)$.

(a) Compute the MLE for the graphical model associated to

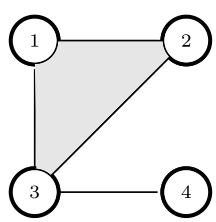


(b) Compute the MLE for the graphical model associated to



(c) What are the ML degrees of the models in (a) and (b)?

5. Consider the simplicial complex Γ below and the associated log-linear model $\mathcal{M}_\Gamma$ of distributions associated to four binary random variables $X = (X_1, X_2, X_3, X_4)$.



Suppose we are given data

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & \end{bmatrix}$$

$$\begin{bmatrix} & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \cdots & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

where each column is a realization (x_1, x_2, x_3, x_4) of X. Find the maximum likelihood estimate $\hat{u}$ of the frequencies u in the model $\mathcal{M}_\Gamma$.

6. Derive the formula for the maximum likelihood estimate of the discrete Markov chain model $\mathcal{M}(I_3)$ from Lecture 1.

7. The following exercise aims to show the proposition regarding the saturation of ideals from the lecture on March 5. You are allowed and encouraged to use all statements presented in the lecture on January 29 (the relevant statements can be found on the first 5 slides).

   (a) Let $Z \subseteq \mathbb{C}^m$ be any subset. Show that $I(Z) = I(V(I(Z)))$.
   (Recall that $V(I(Z))$ is the Zariski closure of Z.)

   A Zariski closed subset of $\mathbb{C}^m$ is called *irreducible* if it cannot be written as the union of two proper Zariski closed subsets.

   Let $I, J \subseteq \mathbb{C}[x_1, \ldots, x_m]$ be ideals. We consider the *irreducible decomposition* of the variety
   $$V(I) = V_1 \cup \ldots \cup V_r,$$
   i.e. each component V_i is an irreducible Zariski closed subset of $\mathbb{C}^m$ and $V_i \nsubseteq V_j$ for $i \neq j$. We may assume that $V_i \nsubseteq V(J)$ for $i \leq k$ and $V_i \subseteq V(J)$ for $i > k$.

   (b) Show that $V(I(V_i \setminus V(J))) = V_i$ for $i \leq k$.
   Deduce that the Zariski closure of $V(I) \setminus V(J)$ is $V_1 \cup \ldots \cup V_k$.
   (You may use the following statement: If $Z \subseteq \mathbb{C}^m$ is an irreducible Zariski closed set and $U \subseteq Z$ is a non-empty Zariski open subset of Z, then U is Zariski dense in Z, i.e. the Zariski closure of U is Z.)

   (c) Show that the saturation ideal $I : J^\infty$ is contained in $I(V_1) \cap \ldots \cap I(V_k)$.
   (Hint: Show that $I : J^\infty \subseteq I(V_i \setminus V(J))$ for all $i \leq k$, and then deduce the statement using (a) and (b).)

   (d) Show that $I(V_1) \cap \ldots \cap I(V_k) \subseteq \sqrt{I : J^\infty}$.

   (e) Show that the Zariski closure of $V(I) \setminus V(J)$ is $V(I : J^\infty)$.
   (Hint: Use (b), (c) and (d).)

8. The following exercise is meant to be done with `Macaulay2` or similar software. An online version of `Macaulay2`, with a short self-contained tutorial, can be found here:

   `https://www.unimelb-macaulay2.cloud.edu.au/#editor`

   Consider the lattice $\mathcal{L} = \ker_{\mathbb{Z}}(A)$ corresponding to the matrix

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

   (a) Compute a lattice basis of $\mathcal{L}$.

   (b) Compute a Markov basis of $\mathcal{L}$.
   (Hint: Use the method `gfanLatticeIdeal` from the package `gfanInterface`. You need to load the package first: `loadPackage "gfanInterface"`.)

   (c) Compute a minimal Markov basis of $\mathcal{L}$. Which degrees do the corresponding generators of the lattice ideal $I_{\mathcal{L}}$ have?
   (Hint: You can apply `mingens` to the lattice ideal $I_{\mathcal{L}}$.)

   (d) Read the short Gröbner basis introduction here:

   `http://www2.macaulay2.com/Macaulay2/doc/Macaulay2-1.12/share/doc/Macaulay2/Macaulay2Doc/html/_what_spis_spa_sp__Groebner_spbasis_qu.html`

   (e) Compute the reduced Gröbner basis of $\mathcal{L}$ with respect to the graded reverse lexicographical monomial order. Which degrees do the corresponding generators of the lattice ideal $I_{\mathcal{L}}$ have?
   (Hint: Make a new ring R with the desired monomial order and transfer the lattice ideal $I_{\mathcal{L}}$ to the new ring using `sub(`$I_{\mathcal{L}}, R$`)`.)