Combinatorial and Algebraic Statistics

Problem Set 2

Due Date: 31 May 2021

1. (a) For $\lambda > 0$, a random variable X has a Poisson distribution with parameter $\lambda > 0$ if its state space is $\mathbb{Z}_{\geq 0}$ and

$$P_\lambda(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad \text{for } k \in \mathbb{Z}_{\geq 0}.$$

Show that the family of Poisson random variables is a regular exponential family. What is its natural parameter space?

(b) Let $G = ([m], E)$ be a graph with m nodes and edge set $E \subset [m] \times [m]$. We denote the adjacency matrix of G by $A = [a_{i,j}] \in \mathbb{R}^{m \times m}$ where $a_{i,j} = 1$ if $(i,j) \in E$ and $a_{i,j} = 0$ if $(i,j) \notin E$. The *Erdös-Renyi graph model* is the one parameter model with state space $\mathbb{G}_m$ consisting of all graphs G on m nodes where the probability

$$P_\theta(G) = \prod_{i \in [m]} \prod_{j \in [m]} \theta^{a_{i,j}}(1-\theta)^{1-a_{i,j}} \qquad \text{where } \theta \in (0,1).$$

Show that the Erdös-Renyi graph model is an exponential family. What is its natural parameter space?

2. Consider the vector $h = (1,1,1,2,2,2)$ and the matrix

$$A = \begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 & 1 & 1 \end{pmatrix}$$

(a) What are the generators of I_A and $I_{A,h}$?

(b) What familiar statistical model is the discrete exponential family $\mathcal{M}_{A,h}$?

3. Suppose we have three particles P_1, P_2, P_3 that each can be in one of two states $+1$ or -1. The state space of our model is pairs of particles and states $\{(P_i, \pm 1) : i \in [3]\}$ (at any one moment a particle appears and we can measure its state), and our data-generating distribution is assumed to be an exponential family with canonical sufficient statistic

$$T(P_i, \pm 1) = (\pm e_i, 1)^t \in \mathbb{R}^4,$$

where e_1, e_2, e_3 denote the standard basis vectors in $\mathbb{R}^3$. Suppose we observe the data

$$\begin{aligned}\mathbb{D} = \{&(P_3,1),(P_1,-1),(P_2,-1),(P_2,1),(P_2,-1),(P_1,-1),(P_2,1),(P_2,1),(P_2,1),(P_1,1),\\ &(P_2,-1),(P_1,-1),(P_2,1),(P_3,1),(P_2,1),(P_3,1),(P_3,1),(P_1,-1),(P_2,-1),(P_3,-1)\}.\end{aligned}$$

Does the MLE of this data exist? If so, what is it?

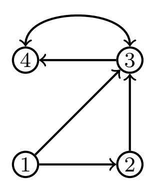
4. Let $(X_1, \ldots, X_m)$ be jointly distributed discrete random variables, and let $G = ([m], E)$ be a DAG in which we assign to each edge $(i,j) \in E$ a subset $\ell(i,j)$ of the outcomes of $X_{\mathrm{pa}_G(j)\setminus\{i\}}$. Given $S \subset [m]$ and an outcome x_S of X_S, we define the DAG G_{x_S} to be the DAG in which we delete all edges (i,j) of G whose label contains an outcome $y_{S\cap(\mathrm{pa}_G(j)\setminus\{i\})}$ of $X_{S\cap(\mathrm{pa}_G(j)\setminus\{i\})}$ all of whose entries coincide with the corresponding entries in x_S. For disjoint subsets $A, B, C, S \subset [m]$ and x_S an outcome of X_S, we say that A and B are *c-separated* in G given C and x_S whenever A and B are d-separated given $C \cup S$ in G_{x_S}. A distribution $\mathbb{P}$ is *Markov* to G if it entails $X_A \perp\!\!\!\perp X_B | X_C, X_S = x_S$ whenever A and B are c-separated given C and x_s in G. Note here that a distribution $P(X_1, \ldots, X_m)$ entailing the relation $X_A \perp\!\!\!\perp X_B | X_C, X_S = x_S$ is equivalent to the condition that

$$P(x_A, x_B | x_C, x_S) = P(x_A | x_C, x_S) P(x_B | x_C, x_S),$$

for all outcomes x_A, x_B, and x_C.

Show that c-separation is not complete: Show there exist DAGs G with edge assignments $\ell(i,j)$ for which any positive distribution Markov to G must entail a relation $X_A \perp\!\!\!\perp X_B | X_C, X_S = x_S$, but A and B are not c-separated given C and x_S in G.

5. Let G be the following mixed graph, and let $\mathcal{M}(G) \subseteq PD_4$ denote its linear structural equation model.



(a) The definition of d-separation can be extended directly to mixed graphs from DAGs. Show that the mixed graph G encodes no d-separation statements, and hence, d-separation captures no constraining equations on the model $\mathcal{M}(G)$.

(b) Show that any $\Sigma = [\sigma_{i,j}] \in \mathcal{M}(G)$ satisfies the constraint

$$\det(\Sigma_{12,34}) = 0,$$

where $\Sigma_{12,34}$ is the submatrix of Σ given by the rows $\{1,2\}$ and the columns $\{3,4\}$. What do these two results combine to say?

6. Consider the multivariate normal model associated to the DAG

Given the data $\mathbb{D} = \{(1,1,0),\ (0,0,1)\}$, find all MLEs.