READING GROUP IN ALGEBRAIC STATISTICS

ELIANA DUARTE AND BEN HOLLERING

For some of these exercises it is desirable to use a computer algebra system. We recommend using the Graphical Models and 4TiTwO packages in Macaulay2. But others are possible. These exercises are a guide, if you saw an exercise that you liked or would like to get deeper into some aspects in the previous chapters, let us know.

Exercise 1. Choose one of the packages `GraphicalModels`, `GraphicalModelsMLE`, `4TiTwo`, or `PhylogeneticTrees` for Algebraic Statistics written in Macaulay2. Use 10 minutes to give an intro to your chosen package and what it can do.

Exercise 2 (Ex. 4.1. Part 1. [2])**.** The conditional independence implication

$$A \perp\!\!\!\perp B|c \cup D \text{ and } A \perp\!\!\!\perp B|D \Rightarrow A \perp\!\!\!\perp (B \cup c)|D \text{ or } (A \cup c) \perp\!\!\!\perp B|D$$

is called the Gaussoid axiom, where c denotes a singleton. Show that if X is a jointly normal random variable, it satisfies the Gaussoid axiom.

Exercise 3 (Ex. 4.7 [2])**.** For four binary random variables, consider the conditional independence model $\mathcal{C} = \{1 \perp\!\!\!\perp 3 \mid \{2,4\}, 2 \perp\!\!\!\perp 4|\{1,3\}\}$. Compute the primary decomposition of $I_{\mathcal{C}}$ and describe the components.

Exercise 4 (Ex. 4.9 [2])**.** In the marginal independence model $C = \{1 \perp\!\!\!\perp 2, 1 \perp\!\!\!\perp 3, 2 \perp\!\!\!\perp 3\}$ from Theorem 4.3.5, find the generators of the ideal of the unique component of I_C that intersects the probability simplex.

Exercise 5. Present the proof of Proposition 4.1.9 about conditional independence. It is a good exercise on marginal distributions of normal random variables and Schur complements.

Exercise 6 (Ex 6.2 [2])**.** Consider the vector $h = (1, 1, 1, 2, 2, 2)$ and the matrix

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

1. Compute generators for the toric ideals I_A and $I_{A,h}$.
2. What familiar statistical model is the discrete exponential family $\mathcal{M}_{A,h}$.

Exercise 7. Consider the linear space L of 4×4 symmetric matrices defined by

$$L = \{K \in \mathbb{R}^{10} : k_{11} + k_{12} + k_{13} = k_{23} + k_{34} = 0\}.$$

Determine generators of the vanishing ideal $I(L^{-1}) \subseteq \mathbb{R}[\sigma]$.

Exercise 8 (Ex 7.3 [2])**.** Consider the log-linear model described parametrically via the formula

$$p_{ijkl} = \frac{1}{Z}\alpha_{ij}\beta_{jk}\gamma_{kl}$$

for discrete random variables X_1, X_2, X_3, X_4 each with two outcomes.

1. Show that this model has ML-degree 1.

Date: May 4, 2023.

2. Realize this model via a Horn uniformization.

Exercise 9 (Ex 13.9 [2])**.** Show that the maximum likelihood degree of any Bayesian network is 1. In particular, explain how to determine rational formulas for the maximum likelihood estimates of parameters in graphical models with directed acyclic graphs. You might also want to checkout Lauritzen's book [4]

Exercise 10. Let G be a DAG. What is the hierarchy of containments of the CI ideals $I_{\text{pairwise}(G)}, I_{\text{local}(G)}, I_{\text{global}(G)}$ in the discrete case? Can you give examples of when the containments are strict? What is known about the Gaussian case? Useful references: [1], [3]

References

1. L. D. Garcia-Puente, M. Stillman and B. Sturmfels. Algebraic Geometry of Bayesian Networks. *Journal of Symbolic Computation*, Special issue on the occasion of MEGA 2003, 39,3,331-355,2005.
2. S. Sullivant. *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.
3. T. Kahle, J. Rauh, S. Sullivant. Algebraic Aspects of Conditional Independence and Graphical Models, Chapter 3 of *Handbook of Graphical Models*, 2017.
4. S. L. Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany
Email address: `eliana.duarte@fc.up.pt`
Email address: `benjamin.hollering@mis.mpg.de`