# Matrix Completion for the Independence Model

Kaie Kubjas, Zvi Rosen

March 17, 2014

Completion of independence models is a problem that arises in algebraic statistics. A pair of discrete random variables $X, Y$ is hypothesized to be independent, but only a subset of the values $P(X = i, Y = j) = p_{ij}$ is known. Under what conditions can a given subset be completed to probabilities satisfying the independence criterion?

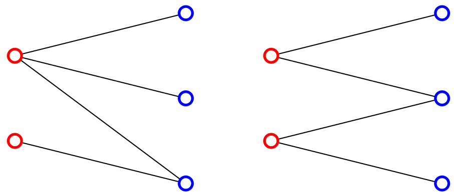
This problem can be rephrased in the language of geometry. Let $\Delta^k$ denote the standard $k$-simplex in $\mathbb{R}^{k+1}$, given by the convex hull of the positive unit vectors in each direction. Define $\Delta_0^k$ as the convex hull of the $k$-simplex and the origin. Let $\mathcal{V}$ be the variety defining rank-1 matrices in $\mathbb{R}^{m\times n}$, and let $\mathcal{V}^* = \mathcal{V} \cap \Delta^{mn-1}$ refer to its intersection with the $mn-1$-simplex. Our question now becomes what are the defining equations and inequalities for different coordinate projections of $\mathcal{V}^*$ into $\Delta_0^k$.

This research lies at the intersection of real algebraic geometry, algebraic statistics and matroid theory. Real algebraic geometry and algebraic statistics are natural parts of the problem as already described. Matroid theory enters the picture because the nature of fibers under coordinate projections is intrinsically linked to the algebraic matroid with ground set the ambient coordinates, and relations defined by the prime ideal associated to the variety. Completion of low-rank matrices with no other constraints was studied in [2] and [1], and used algebraic matroid theory as the setting for completion techniques. Those techniques will be used and adapted as necessary for these applications. We include a quick example to demonstrate how an algebraic matroid arises in this problem.

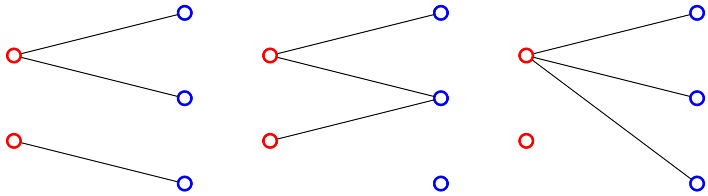
**Example 1.** *Suppose we have a* $2 \times 3$ *matrix of rank* 1*. The ideal is defined by the* $2 \times 2$ *minors of the matrix*

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}$$

*and the linear relation* $x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} = 1$*. The bases of the rank-*1 *determinantal matroid correspond to spanning trees of the bipartite graph* $K_{2,3}$ *([2, Prop 2.6.45]). In our example, the graphs corresponding to bases of the determinantal matroid are:*



*Removing each edge from each potential base, we obtain three bases of the probability matroid*



*The corresponding inequalities they must satisfy to be completable to rank* 1 *are:*

$$
\begin{aligned}
(1-a_1-a_2-b_3)^2-4b_3(a_1+a_2) &\geq 0,\\
a_2(1-a_1-a_2-b_2)-a_1b_2 &\geq 0,\\
&\varnothing,
\end{aligned}
$$

*respectively; in addition to the inequalities defining the* 2*-simplex.*

We aim to explore several conjectures regarding these probability completion matroids:

1. For all $m \times n$, the set of bases can be formed by deleting an edge from any base of the determinantal matroid.
2. For any base, some quadratic inequality defines the image of the projection.
3. Similar statements may hold for higher rank probability matrices, i.e. not independent probabilities but independent conditional on hidden variables.

## References


[1] Franz Király, Zvi Rosen, and Louis Theran. Algebraic matroids with graph symmetry. Unpublished manuscript. http://arxiv.org/abs/1312.3777, 2014.

[2] Franz J. Király, Louis Theran, Ryota Tomioka, and Takeaki Uno. The algebraic combinatorial approach for low-rank matrix completion. *Pre-print.*, 2012. http://arxiv.org/abs/1211.4116.