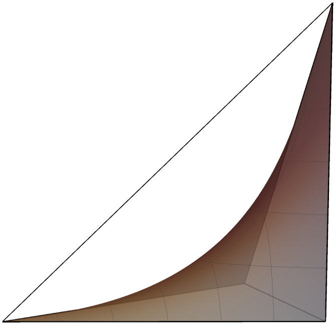
Marginal independence models

Tobias Boege[1] Sonja Petrović[2] Bernd Sturmfels[3,4]

[1]Aalto University [2]Illinois Tech [3]UC Berkeley [4]MPI-MiS

Applied CATS Seminar,
KTH Stockholm, 14 February 2023

Statistical models are semialgebraic sets[1]

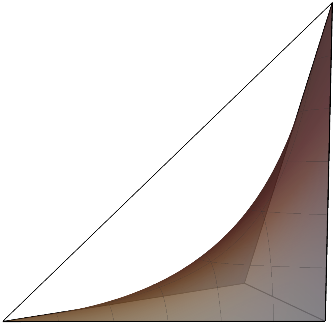


The set of all distributions of two *independent* binary random variables (X, Y) is a surface in the probability simplex defined by

$$P(X = 0, Y = 0) \cdot P(X = 1, Y = 1) = P(X = 0, Y = 1) \cdot P(X = 1, Y = 0).$$

[1] sometimes

Statistical models are semialgebraic sets[1]



The set of all distributions of two *independent* binary random variables (X, Y) is a surface in the probability simplex defined by

$$p_{00} \cdot p_{11} = p_{01} \cdot p_{10}.$$

Also known as the Segre embedding of $\mathbb{P}^1 \times \mathbb{P}^1$.

[1] sometimes

Setup

- Consider discrete random variables X_j with state space $[d_j] = \{1, \ldots, d_j\}$.
- A probability distribution P is identified with the $d_1 \times \cdots \times d_n$ tensor of atomic probabilities $p_{i_1 \ldots i_n} \coloneqq P(X_1 = i_1, \ldots, X_n = i_n)$.
- The probability simplex is the set of all discrete distributions

$$\Delta = \Delta(d_1, d_2, \ldots, d_n) = \{ P \in \mathbb{R}^{d_1 \times \cdots \times d_n} : P \geq 0 \text{ and } \sum P = 1 \}.$$

Setup

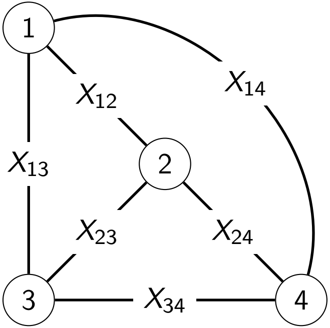
- Consider discrete random variables X_j with state space $[d_j] = \{1, \dots, d_j\}$.
- A probability distribution P is identified with the $d_1 \times \cdots \times d_n$ tensor of atomic probabilities $p_{i_1 \dots i_n} \coloneqq P(X_1 = i_1, \dots, X_n = i_n)$.
- The probability simplex is the set of all discrete distributions

$$\Delta = \Delta(d_1, d_2, \dots, d_n) = \{P \in \mathbb{R}^{d_1 \times \cdots \times d_n} : P \geq 0 \text{ and } \sum P = 1\}.$$

- A statistical model is a subset of Δ. E.g., the binary independence model is the set of all 2×2 matrices $P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$ in $\Delta(2,2)$ such that $\det P = 0$.

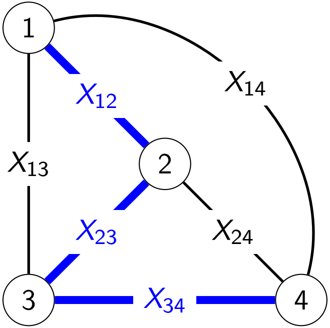
A random graph model

The binary random variables $(X_e)_{e \in E(G)}$ pick a random subgraph such that appearances of edges which do not contain a cycle are completely independent.



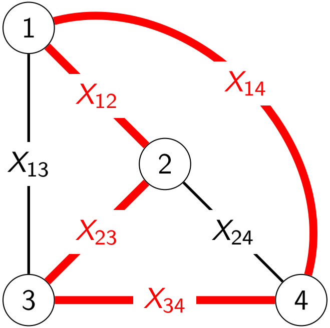
A random graph model

The binary random variables $(X_e)_{e \in E(G)}$ pick a random subgraph such that appearances of edges which do not contain a cycle are completely independent.



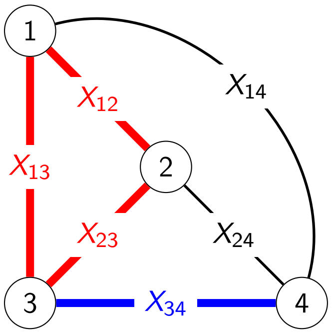
A random graph model

The binary random variables $(X_e)_{e \in E(G)}$ pick a random subgraph such that appearances of edges which do not contain a cycle are completely independent.



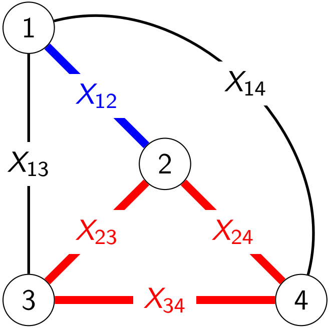
A random graph model

The binary random variables $(X_e)_{e \in E(G)}$ pick a random subgraph such that appearances of edges which do not contain a cycle are completely independent.



A random graph model

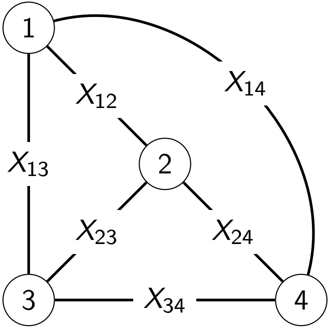
The binary random variables $(X_e)_{e \in E(G)}$ pick a random subgraph such that appearances of edges which do not contain a cycle are completely independent.



A random graph model

The binary random variables $(X_e)_{e \in E(G)}$ pick a random subgraph such that appearances of edges which do not contain a cycle are completely independent.

This describes a statistical model in $\Delta(2, 2, \ldots, 2)$. A point in the model is a probability distribution whose outcomes are graphs on four vertices.



Marginal independence models: Definition

In this talk, a *simplicial complex* is a collection Σ of subsets of $[n]$ such that:

- $\{i\} \in \Sigma$ for all $i \in [n]$,
- $\tau \subseteq \sigma \in \Sigma \Rightarrow \tau \in \Sigma$.

Marginal independence models: Definition

In this talk, a *simplicial complex* is a collection Σ of subsets of $[n]$ such that:

- $\{i\} \in \Sigma$ for all $i \in [n]$,
- $\tau \subseteq \sigma \in \Sigma \Rightarrow \tau \in \Sigma$.

Definition

The *marginal independence model* $\mathcal{M}_\Sigma$ is the set of distributions of $(X_1, \dots, X_n)$ in $\Delta(d_1, \dots, d_n)$ such that X_σ is completely independent for all $\sigma \in \Sigma$.

- The random subgraph model is a marginal independence model where Σ is the simplicial complex of all forests in the graph.

Marginal independence models: Algebra

A subvector X_σ, $\sigma \subseteq [n]$, is *completely independent* if for all choices $i_j \in [d_j]$:

$$P(X_j = i_j : j \in \sigma) = \prod_{j \in \sigma} P(X_j = i_j).$$

That is, the marginal distribution P_σ of X_σ is a tensor of rank 1.

Marginal independence models: Algebra

A subvector X_σ, $\sigma \subseteq [n]$, is *completely independent* if for all choices $i_j \in [d_j]$:

$$P(X_j = i_j : j \in \sigma) = \prod_{j \in \sigma} P(X_j = i_j).$$

That is, the marginal distribution P_σ of X_σ is a tensor of rank 1.

Implicitization of the above parametrization gives the equations of the *Segre variety* $\times_{j \in \sigma} \mathbb{P}^{d_j - 1}$ in $\mathbb{P}^{\prod_{j \in \sigma} d_j - 1}$.

Not to be confused with: Hierarchical models

Hierarchical models are also derived from simplicial complexes but their parametrization is:

$$P(X_j = i_j : j \in [n]) = \prod_{\sigma \text{ facet of } \Sigma} \theta^{(\sigma)}_{i_\sigma}.$$

- Parametrization is for the entire tensor instead of marginals.
- One set of parameters per facet instead of faces factorizing.

Example: $\Sigma = [12, 13, 23]$

- The hierarchical model is known as the "no 3-way interaction model"

$$p_{ijk} = \theta^{(12)}_{ij}\theta^{(13)}_{ik}\theta^{(23)}_{jk}.$$

  For binary variables, its complex variety has dimension 7 and degree 4.
  It is cut out by the quartic $p_{000}p_{011}p_{101}p_{110} - p_{001}p_{010}p_{100}p_{111}$.

- The marginal independence is given implicitly by factorizations of marginal distributions

$$\sum_k p_{ijk} = \sum_{j,k} p_{ijk} \cdot \sum_{i,k} p_{ijk}, \quad \sum_j p_{ijk} = \sum_{j,k} p_{ijk} \cdot \sum_{i,j} p_{ijk}, \quad \sum_i p_{ijk} = \sum_{i,k} p_{ijk} \cdot \sum_{i,j} p_{ijk}.$$

  Its dimension is 5 and it has degree 8.

Kirkup's parametrization

Lemma (Kirkup (2007))

The marginal independence model equals $\mathcal{M}_\Sigma = \mathcal{S} + \mathcal{L}_\Sigma$ where $\mathcal{L}_\Sigma$ is the linear subspace with marginals $P_\sigma = 0$ for all $\sigma \in \Sigma$.

Proof.

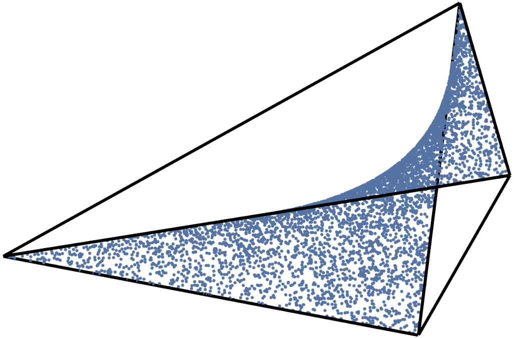
- Given $P \in \mathcal{M}_\Sigma$, take its marginals P_j, $j \in [n]$, corresponding to the distributions of the individual random variables X_j.
- $P' = \bigotimes_j P_j \in \mathcal{S}$ and $P - P' \in \mathcal{L}_\Sigma$ since P and P' have identical marginals and P_σ and P'_σ are both completely independent.

Note: This parametrization is just $P = \mathrm{MLE}_{\mathcal{S}}(P) + \text{correction term}$.

Kirkup's parametrization

Lemma (Kirkup (2007))

The marginal independence model equals $\mathcal{M}_\Sigma = \mathcal{S} + \mathcal{L}_\Sigma$ *where* $\mathcal{L}_\Sigma$ *is the linear subspace with marginals* $P_\sigma = 0$ *for all* $\sigma \in \Sigma$*.*



Möbius coordinates

The defining ideal of $\mathcal{M}_\Sigma$ is generated by homogeneous, quadratic polynomials coming from the Segre equations for each $\sigma \in \Sigma$, e.g., for $\Sigma = [12, 13, 23]$,

$$p_{000}p_{110} + p_{001}p_{110} + p_{000}p_{111} + p_{001}p_{111} = p_{010}p_{100} + p_{011}p_{100} + p_{010}p_{101} + p_{011}p_{101} \qquad (1 \perp\!\!\!\perp 2)$$

$$p_{000}p_{101} + p_{010}p_{101} + p_{000}p_{111} + p_{010}p_{111} = p_{001}p_{100} + p_{011}p_{100} + p_{001}p_{110} + p_{011}p_{110} \qquad (1 \perp\!\!\!\perp 3)$$

$$p_{000}p_{011} + p_{011}p_{100} + p_{000}p_{111} + p_{100}p_{111} = p_{001}p_{010} + p_{010}p_{101} + p_{001}p_{110} + p_{101}p_{110} \qquad (2 \perp\!\!\!\perp 3)$$

Möbius coordinates

The defining ideal of $\mathcal{M}_\Sigma$ is generated by homogeneous, quadratic polynomials coming from the Segre equations for each $\sigma \in \Sigma$, e.g., for $\Sigma = [12, 13, 23]$,

$$q_\varnothing q_{12} = q_1 q_2 \qquad (1 \perp\!\!\!\perp 2)$$
$$q_\varnothing q_{13} = q_1 q_3 \qquad (1 \perp\!\!\!\perp 3)$$
$$q_\varnothing q_{23} = q_2 q_3 \qquad (2 \perp\!\!\!\perp 3)$$

In the *Möbius coordinates* $q_\bullet$, the ideal becomes toric.

Toric representation theorem

Theorem

The variety of the marginal independence model $\mathcal{M}_\Sigma$ is irreducible and its prime ideal is toric in Möbius coordinates. That is, it has a parametrization by monomials and its ideal is generated by binomials. The parametrization is

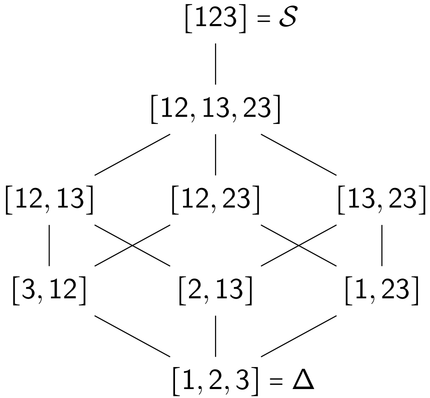
$$q_{i_1\dots i_n} \mapsto \prod_{j:i_j\neq +} \theta_{i_j}^{(j)} \quad \text{for} \quad \{j : i_j \neq +\} \in \Sigma.$$

Moreover, the statistical model $\mathcal{M}_\Sigma$ is a contractible semialgebraic set of dimension

$$\sum_{j=1}^{n}(d_j - 1) + \sum_{\tau\notin\Sigma}\prod_{j\in\tau}(d_j - 1).$$

Marginal independence models: Properties

- Nice parametrization as Segre + linear space.
- Nice binomial equations in Möbius coordinates (but degrees can be high).
- Contractible statistical models.
- Stratify the probability simplex.
- Contain our random graph models and more!



Better coordinates for conditional independence ideals

Consider the constraints $\{X_1 \perp\!\!\!\perp X_2, X_1 \perp\!\!\!\perp X_2 \mid (X_3, X_4), X_1 \perp\!\!\!\perp X_4, X_2 \perp\!\!\!\perp X_4, X_3 \perp\!\!\!\perp X_4\}$ on four binary random variables. Does there exist a distribution which satisfies all of them and no others?

$$p_{0100}p_{1000} = p_{0000}p_{1100}, \quad p_{0101}p_{1001} = p_{0001}p_{1101}, \quad p_{0110}p_{1010} = p_{0010}p_{1110}, \quad p_{0111}p_{1011} = p_{0011}p_{1111}$$

$$p_{0100}p_{1000} + p_{0101}p_{1000} + p_{0110}p_{1000} + p_{0111}p_{1000} + p_{0100}p_{1001} + p_{0101}p_{1001} + p_{0110}p_{1001} + p_{0111}p_{1001} +$$
$$p_{0100}p_{1010} + p_{0101}p_{1010} + p_{0110}p_{1010} + p_{0111}p_{1010} + p_{0100}p_{1011} + p_{0101}p_{1011} + p_{0110}p_{1011} + p_{0111}p_{1011} =$$
$$p_{0000}p_{1100} + p_{0001}p_{1100} + p_{0010}p_{1100} + p_{0011}p_{1100} + p_{0000}p_{1101} + p_{0001}p_{1101} + p_{0010}p_{1101} + p_{0011}p_{1101} +$$
$$p_{0000}p_{1110} + p_{0001}p_{1110} + p_{0010}p_{1110} + p_{0011}p_{1110} + p_{0000}p_{1111} + p_{0001}p_{1111} + p_{0010}p_{1111} + p_{0011}p_{1111}$$

$$p_{0001}p_{1000} + p_{0011}p_{1000} + p_{0101}p_{1000} + p_{0111}p_{1000} + p_{0001}p_{1010} + p_{0011}p_{1010} + p_{0101}p_{1010} + p_{0111}p_{1010} +$$
$$p_{0001}p_{1100} + p_{0011}p_{1100} + p_{0101}p_{1100} + p_{0111}p_{1100} + p_{0001}p_{1110} + p_{0011}p_{1110} + p_{0101}p_{1110} + p_{0111}p_{1110} =$$
$$p_{0000}p_{1001} + p_{0010}p_{1001} + p_{0100}p_{1001} + p_{0110}p_{1001} + p_{0000}p_{1011} + p_{0010}p_{1011} + p_{0100}p_{1011} + p_{0110}p_{1011} +$$
$$p_{0000}p_{1101} + p_{0010}p_{1101} + p_{0100}p_{1101} + p_{0110}p_{1101} + p_{0000}p_{1111} + p_{0010}p_{1111} + p_{0100}p_{1111} + p_{0110}p_{1111}$$

....

Better coordinates for conditional independence ideals

Consider the constraints $\{X_1 \perp\!\!\!\perp X_2, X_1 \perp\!\!\!\perp X_2 \mid (X_3, X_4), X_1 \perp\!\!\!\perp X_4, X_2 \perp\!\!\!\perp X_4, X_3 \perp\!\!\!\perp X_4\}$ on four binary random variables. Does there exist a distribution which satisfies all of them and no others?

$$q_3 q_4 q_{1234} = q_{134} q_{234}$$

$$q_3 q_{123} q_4 + q_{13} q_{23} + q_{134} q_{234} + q_3 q_{1234} = q_3 q_4 q_{1234} + q_3 q_{123} + q_{23} q_{134} + q_{13} q_{234}$$

$$q_1 q_2 q_4^2 + q_3 q_4 q_{124} + q_{134} q_{234} + q_4 q_{1234} = q_2 q_4 q_{134} + q_1 q_4 q_{234} + q_3 q_4 q_{1234} + q_4 q_{124}$$

$$q_1 q_2 q_4^2 + q_1 q_2 q_3 + q_2 q_{13} q_4 + q_1 q_{23} q_4 + q_3 q_{123} q_4 + q_3 q_4 q_{124} + q_{13} q_{23} + q_2 q_{134} +$$
$$q_1 q_{234} + q_{134} q_{234} + q_3 q_{1234} + q_4 q_{1234} + q_{123} + q_{124} =$$
$$q_1 q_2 q_3 q_4 + q_1 q_2 q_4 + q_2 q_4 q_{134} + q_1 q_4 q_{234} + q_3 q_4 q_{1234} + q_2 q_{13} + q_1 q_{23} + q_3 q_{123} +$$
$$q_{123} q_4 + q_3 q_{124} + q_4 q_{124} + q_{23} q_{134} + q_{13} q_{234} + q_{1234}.$$

Better coordinates for conditional independence ideals

Consider the constraints $\{X_1 \perp\!\!\!\perp X_2, X_1 \perp\!\!\!\perp X_2 \mid (X_3, X_4), X_1 \perp\!\!\!\perp X_4, X_2 \perp\!\!\!\perp X_4, X_3 \perp\!\!\!\perp X_4\}$ on four binary random variables. Does there exist a distribution which satisfies all of them and no others? Yes!

$$q_{1234} = \frac{q_{134} q_{234}}{q_3 q_4}$$

$$q_{123} = \frac{q_4 q_{13} q_{23} - q_4 q_{13} q_{234} - q_4 q_{134} q_{23} + q_{134} q_{234}}{q_3 q_4 (1 - q_4)}$$

$$q_{124} = \frac{q_{134} q_{234} - q_{134} q_2 q_3 q_4 - q_1 q_{234} q_3 q_4 + q_1 q_2 q_3 q_4^2}{q_3 q_4 (1 - q_3)}$$

$$q_{134} = \frac{q_{13}((q_{234} q_4 - q_2 q_3 q_4^2) - (q_{23} q_4 - q_2 q_3 q_4)) + q_1 q_3 q_4 (1 - q_4)(q_{23} - q_2 q_3 (q_{234} - q_{23} q_4))}{q_{234} - q_{23} q_4}$$

Parameter estimation

Given a statistical model $\mathcal{M}$ and a sample distribution $U \in \Delta$, we seek the point in $\mathcal{M}$ which best "explains" the observations in U.

Parameter estimation

Given a statistical model $\mathcal{M}$ and a sample distribution $U \in \Delta$, we seek the point in $\mathcal{M}$ which best "explains" the observations in U.

- Maximum likelihood: $\max \sum u_\bullet \log p_\bullet$ s.t. $P \in \mathcal{M}$.
- Euclidean distance: $\min \sum \| u_\bullet - p_\bullet \|^2$ s.t. $P \in \mathcal{M}$.

Parameter estimation

Given a statistical model $\mathcal{M}$ and a sample distribution $U \in \Delta$, we seek the point in $\mathcal{M}$ which best "explains" the observations in U.

- Maximum likelihood: $\max \sum u_\bullet \log p_\bullet$ s.t. $P \in \mathcal{M}$.
- Euclidean distance: $\min \sum \| u_\bullet - p_\bullet \|^2$ s.t. $P \in \mathcal{M}$.

For $\mathcal{M} = \mathcal{S}(2,2,2)$, i.e., $\Sigma = [123]$, and $U = (2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}, \underline{2^{-7}})$:

	Deg	# Real	$\hat{p}_{000}$	$\hat{p}_{001}$	$\hat{p}_{010}$	$\hat{p}_{011}$	$\hat{p}_{100}$	$\hat{p}_{101}$	$\hat{p}_{110}$	$\hat{p}_{111}$
ED	17	1	0.500	0.250	0.125	0.062	0.032	0.016	0.008	0.004
ML	1	1	0.496	0.250	0.126	0.063	0.033	0.016	0.008	0.004

Computed using `HomotopyContinuation.jl`.

Database of small models

https://mathrepo.mis.mpg.de/MarginalIndependence

dimension	degree	mingens	f-vector	simplicial complex Σ	ED	ML
15	1	()	$(1,4)_5$	$[1,2,3,4]$	1	1
14	2	(1)	$(1,4,1)_6$	$[3,4,12]$	5	1
13	3	(3)	$(1,4,2)_7$	$[4,12,13]$	5	9
13	4	(2)	$(1,4,2)_7$	$[14,23]$	25	1041
12	4	(6)	$(1,4,3)_8$	$[12,13,14]$	5	209
12	5	(5)	$(1,4,3)_8$	$[12,14,23]$	21	1081
12	5	(5)	$(1,4,3)_8$	$[4,12,13,23]$	21	17
			...			
8	16	(21)	$(1,4,6,1)_{12}$	$[14,24,34,123]$	117	8542
7	18	(28)	$(1,4,6,2)_{13}$	$[34,123,124]$	89	2121
6	20	(36)	$(1,4,6,3)_{14}$	$[123,124,134]$	89	505
5	23	(44)	$(1,4,6,4)_{15}$	$[123,124,134,234]$	169	561
4	24	(55)	$(1,4,6,4,1)_{16}$	$[1234]$	73	1

Open ends

- The CI model of $1 \perp\!\!\!\perp \{2, 3\}$ is not of marginal independence type but nevertheless it is toric in Möbius coordinates: $q_{12} = q_1 q_2$, $q_{13} = q_1 q_3$, $q_{123} = q_1 q_{23}$ ...
- Kirkup: Is the toric variety of $\mathcal{M}_\Sigma$ always Cohen-Macaulay?
- Side story: Entropic matroids.
- Are the open models $\mathcal{M}_\Sigma \cap \Delta^\circ$ smooth manifolds?
- How to select a fitting marginal independence model for given data?
- Is the real solution to the affine ED problem generically unique?

References

Tobias Boege, Sonja Petrović, and Bernd Sturmfels. "Marginal Independence Models". In: *Proceedings of the 2022 International Symposium on Symbolic and Algebraic Computation*. ISSAC '22. Villeneuve-d'Ascq, France: Association for Computing Machinery (ACM), 2022, pp. 263–271. DOI: 10.1145/3476446.3536193.

Mathias Drton and Thomas S. Richardson. "Binary models for marginal independence". In: *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 70.2 (2008), pp. 287–309. DOI: 10.1111/j.1467-9868.2007.00636.x.

George A. Kirkup. "Random variables with completely independent subcollections". In: *J. Algebra* 309.2 (2007), pp. 427–454. DOI: 10.1016/j.jalgebra.2006.06.023.

Seth Sullivant. *Algebraic Statistics*. Vol. 194. Graduate Studies in Mathematics. American Mathematical Society, 2018.