BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 58, Number 2, April 2021, Pages 305–309
https://doi.org/10.1090/bull/1727
Article electronically published on February 24, 2021

Algebraic statistics, by Seth Sullivant, Graduate Studies in Mathematics, Vol. 194, American Mathematical Society, Providence, RI, 2018, xiii+409 pp., ISBN 978-1-4704-3517-2

Algebraic statistics studies the interplay of algebra and statistics. Both of these fields are usually broadly interpreted. Algebra can include algebraic geometry, combinatorics, discrete mathematics, optimization, and so on. Statistics is usually meant to include also aspects of probability theory and data science in general. Algebraic statistics has the MSC number 62R01, a dedicated journal *Algebraic statistics*, and a vibrant community. There are several spezialized textbooks on algebraic statistics themes, e.g., [1–3, 6, 8], but Seth Sullivant's text is the first to aim to capture the state of algebraic statistics in its entirety.

Algebraic statistics is not centered around one definition like topology, which studies topological spaces, or group theory, which studies groups. Rather, algebraic statistics is a realm of ideas and a community of researchers who constantly discover new areas of activity connected to algebra and data science. Importantly, the transfer of ideas goes in both directions. Mathematical structures in the applications motivate new fundamental theorems, and algebraic theorems solve statistical problems.

Seth Sullivant's new book is a general introduction and exposition, aiming to take account of the state of the art. After a few introductory chapters with prerequisites, each new chapter goes to a different topic and lays out the foundations of another success story of algebraic statistics. To give a flavor of this, we showcase conditional independence here, which, as Sullivant writes, "is a natural place to begin to study the connection between probabilistic concepts and methods from algebraic geometry".

Conditional independence constraints are used to describe a statistical model, that is, a set of probability distributions in which a true, data-generating distribution is sought. Conditional independence constraints are natural because humans, when doing science, always think about cause and effect, dependence and independence. While discovering actual causality using statistics is another rapidly developing and debated topic [4, 5], here we content ourselves with statistical (in)dependence. To quote again: "conditional independence constraints are simple and intuitive restrictions on probability densities that allow one to express the notion that two sets of random variables are unrelated, typically given knowledge of the values of a third set of random variables".

Chapter 4 of *Algebraic statistics* contains an introduction to conditional independence. Here we illustrate the connection to algebra on distributions of random variables with finitely many discrete values. Another important class with many algebraic structures are normally distributed random variables. In general, (conditional) independence constraints concern factorization properties of joint densities of a collection of random variables, but for these two cases, conditional independence constraints take the form of *polynomial* conditions on the densities. For discrete random variables, this means polynomial conditions on the elementary probabilities of events. In the case of normally distributed random variables, the conditions are rank conditions on submatrices of covariances.

To get something tangible—something three-dimensional—consider two binary random variables X_1 and X_2. For $i, j \in \{0, 1\}$. we denote by

$$p_{ij} = \text{Prob}(X_1 = i, X_2 = j)$$

the probability that X_1 takes the value i and at the same time X_2 takes the value j. This is a 2×2-matrix and, being probabilities, the entries satisfy $p_{ij} \geq 0$ and $\sum_{ij} p_{ij} = 1$. This means that all possible distributions for two binary random variables form a three-dimensional simplex in real 4-space. The condition that X_1 and X_2 are (unconditionally) independent is denoted $X_1 \perp\!\!\!\perp X_2$. It holds if and only if the matrix $(p_{ij})_{ij}$ has rank 1, or respectively, the determinant $p_{00}p_{11} - p_{10}p_{01}$ vanishes. This in turn happens if and only if the small probability matrix factors as a product of two vectors,

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} q_0 \\ q_1 \end{pmatrix} \begin{pmatrix} r_0 & r_1 \end{pmatrix}.$$

This is the factorization of the densities that one often sees in the definition of statistical independence. In this special case of unconditional independence, the two vectors turn out to be the marginal distributions of X_1 and X_2, but this need not hold in general.

Figure 1 shows the surface of distributions that satisfy the independence constraint. This surface is well known. For example, it is a doubly ruled surface that algebraic geometers might recognize as $\mathbb{P}^1 \times \mathbb{P}^1$ in its Segre embedding into $\mathbb{P}^3$.

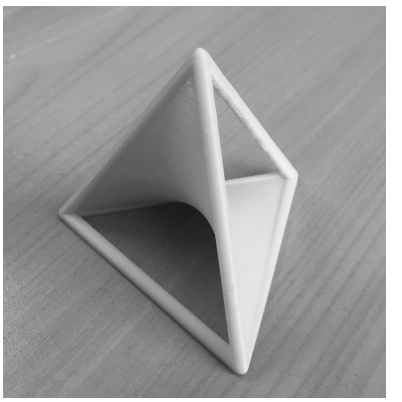

FIGURE 1. The independent binary distributions inside the probability simplex, 3D-printed by the author of this review.

In the general case of n discrete random variables $X_1, \ldots, X_n$, where the kth variable takes r_k states, the distribution is an $r_1 \times \cdots \times r_n$-tensor. If $n \geq 3$, there are enough random variables to condition on something. A conditional independence (CI for short) statement is written $X_A \perp\!\!\!\perp X_B \,|X_C$, where A, B, C are disjoint subsets of the set $\{1, \ldots, n\}$ of indices of the random variables. Then X_A stands for the random variables X_i with $i \in A$, and so on. The CI-statement expresses that knowing the value of X_C explains all dependence between X_A and X_B. For example, in some population, hair loss and the amount of football watched on TV may be statistically dependent, but become independent when conditioned on other factors such as gender. Also for conditional independence among n random variables, quadratic equations in the elementary probabilities appear. It is therefore

plausible that questions about these models are connected to rank conditions and determinant-like equations on tensors, and those have been of interest in commutative algebra, algebraic geometry, and beyond.

Moving to the second large class of distributions, a normally distributed random vector $(Y_1, \ldots, Y_n)$ is specified by a mean $\mu \in \mathbb{R}^n$ and a positive-definite *covariance matrix* $\Sigma \in \mathbb{R}^{n \times n}$. For conditional independence, the mean is irrelevant. Sullivant teaches us that a conditional independence statement $Y_A \perp\!\!\!\perp Y_B \,|Y_C$ specializes to rank conditions on the covariance submatrix $\Sigma_{AC \times BC}$ whose rows are indexed by $A \cup C$ and whose columns are indexed by $B \cup C$. Precisely, the CI holds if and only if the rank of $\Sigma_{AC \times BC}$ is $|C|$, the cardinality of C. Since Σ is positive definite, this rank is at least $|C|$, so it being equal to $|C|$ is equivalent to the vanishing of the $|C|+1$-minors of $\Sigma_{AC \times BC}$, again an algebraic constraint.

One might observe that not everything is strictly algebraic. In both of the above scenarios some nonnegativity entered, and for a complete understanding of the situation algebraic statistics often combines algebraic geometry results (which require the Nullstellensatz and thus algebraically closed fields) with techniques from real algebra. Section 6.4 of Sullivant's book contains an appetizer.

That conditional independence is represented by algebraic constraints might not seem very exciting, before seeing what can be done with it. We look at a first application featured in Example 4.3.2 of the book. In practice it is common to specify a model by requiring several CI constraints. So, consider the set of distributions for three binary random variables X_1, X_2, X_3 that satisfy both $X_1 \perp\!\!\!\perp X_2 \,|X_3$ and $X_2 \perp\!\!\!\perp X_3$. Denoting the distribution as the entries of a $2 \times 2 \times 2$-tensor p_{ijk} as above, it turns out we are looking at distributions that satisfy

$$p_{000}p_{110} = p_{010}p_{100}, \quad p_{001}p_{111} = p_{011}p_{101} \qquad (X_1 \perp\!\!\!\perp X_2 \,|X_3\,), \tag{1}$$

$$(p_{000} + p_{100})(p_{011} + p_{111}) = (p_{001} + p_{101})(p_{010} + p_{110}) \qquad (X_2 \perp\!\!\!\perp X_3\,). \tag{2}$$

This is a system of polynomial equations, and primary decomposition from commutative algebra allows us to solve it. Such a solution is writing the model as a union of finitely many irreducible pieces, thereby turning the AND conditions into OR conditions. In the best possible world (and this example), the irreducible pieces have a nice probabilistic interpretation. Using a computer algebra system (Sullivant's book contains many code examples for different systems), we find that if a distribution p_{ijk} of X_1, X_2, X_3 satisfies the above equations, then either:

(a) all of the following hold,

$$\begin{aligned} &p_{010}p_{100} - p_{000}p_{110} = 0, && p_{101}p_{110} - p_{100}p_{111} = 0, \\ &p_{010}p_{101} - p_{000}p_{111} = 0, && p_{011}p_{100} - p_{001}p_{110} = 0, \\ &p_{001}p_{010} - p_{000}p_{011} = 0, && p_{011}p_{101} - p_{001}p_{111} = 0; \text{ or} \end{aligned} \tag{3}$$

(b) all of the following equations hold,

$$p_{001}p_{111} = p_{011}p_{101}, \qquad p_{000} + p_{100} = 0, \qquad p_{010} + p_{110} = 0; \text{ or} \tag{4}$$

(c) all of the following equations hold,

$$p_{000}p_{110} - p_{010}p_{100}, \qquad p_{011} + p_{111} = 0, \qquad p_{001} + p_{101} = 0. \tag{5}$$

A trained eye can observe that (3) expresses $X_2 \perp\!\!\!\perp \{X_1, X_3\}$, so that X_2 is (unconditionally) independent of X_1 and X_3. And fortunately in this example, the two other possibilities are not relevant for modeling. We examine (5), the case (4)

being similar. Since p_{ijk} are assumed nonnegative (as probabilities), we can conclude from the last two equations of (5) that $p_{011} = p_{111} = p_{001} = p_{101} = 0$. For reasons that become clear momentarily, we now plug these into (3). Leaving out all equations that are implied because both terms vanish already, only one quadratic equation remains:

$$p_{011} = p_{111} = p_{001} = p_{101} = 0, \qquad p_{010}p_{100} - p_{000}p_{110} = 0.$$

Up to an irrelevant sign, this is just (5) again. This means that the nonnegative solutions of (5) are contained in the nonnegative solutions of (3). Thus $X_2 \perp\!\!\!\perp \{X_1, X_3\}$ holds for them, and we need not further consider (5) when analyzing the model. What we have derived here is a conditional independence implication,

$$(\, X_1 \perp\!\!\!\perp X_2 \,|\, X_3 \ \text{ AND } \ X_2 \perp\!\!\!\perp X_3 \,) \implies X_2 \perp\!\!\!\perp \{X_1, X_3\}\,.$$

This example demonstrates how, using algebraic statistics, computer algebra can make logical inference about conditional independence possible. A highlight in this direction is Sullivant's proof that Gaussian conditional independence cannot be finitely axiomatized (Section 4.3.3). A similar result for discrete random variables has been achieved by Studený [7]. Those who wonder why the primary decomposition showed the third possibility as a component should note that (5) does not imply (3) without the nonnegativity assumption. This is in fact a common phenomenon in algebraic statistics: to gain insight, we combine algebraic and semialgebraic constraints.

After this appetizer from the introductory chapters, logicians and commutative algebraists can team up and find more on conditional independence implications in Chapters 4, 9, 13, and 14. From this point forward, the study of conditional independence continues with graphical models. There collections of independence statements are derived from graphs using separation statements. Said differently, interactions among random variables are modeled using edges and connectivity in a graph. This natural representation of interaction goes back to statistical physics, genetics, and the analysis of contingency tables. Algebra has many powerful tools to contribute, and algebraic statistics contributes foundational insight.

The nineteen chapters of *Algebraic statistics* by Seth Sullivant have something to offer everyone. To name a few more, a combinatorialist can enjoy lattice walks in Chapter 9 or random graphs in Chapter 11. An optimization expert might consider integer programming approaches to privacy questions in Chapter 10. The mathematical biologist finds phylogenetics in Chapter 15. A geometer could discover the tropical Grassmannian while wandering from the finite metric spaces in phylogenetics and using Chapter 19 as a map, and so on. Sullivant's book offers an overview of algebraic statistics today and it highlights the many different aspects of the field. Each new chapter is an invitation to explore one area of algebraic statistics further. It provides all the basics for getting started as well as relevant pointers to the literature. This does not imply that the chapters would be completely independent. Coherence comes through several connecting elements that might be called the core of algebraic statistics. Conditional independence was already highlighted. Maximum likelihood estimation with its task to solve critical equations is another.

As a service to the readership, Sullivant included four introductory chapters to the book. These introduce ideas and techniques from algebra and statistics. Using them, the book can serve, and has served, in teaching the next generation of algebraic statisticians. The book also regularly comes back to statistical practice,

realistic models, and datasets from the literature. Useful computer code snippets help to get started with the relevant software systems. Clearly, the author and, more generally, the algebraic statistics community have statistical practice in mind.

Sullivant's new textbook has been longed for as it covers algebraic statistics in greater breadth than any previous text. It is suitable for self-study, courses, and reference. The different chapters can appeal to a wide range of mathematicians, and they have just the right size to spark interest without being overwhelming. This book can only be recommended to everyone interested in this interdisciplinary field.

References

[1] Satoshi Aoki, Hisayuki Hara, and Akimichi Takemura, *Markov bases in algebraic statistics*, Springer Series in Statistics, vol. 199, Springer, New York, 2012. MR2961912

[2] Cristiano Bocci and Luca Chiantini, *An introduction to algebraic statistics with tensors*, Unitext, vol. 118, Springer, Cham, 2019. MR3969980

[3] Mathias Drton, Bernd Sturmfels, and Seth Sullivant, *Lectures on algebraic statistics*, Oberwolfach Seminars, vol. 39, Birkhäuser Verlag, Basel, 2009. MR2723140

[4] Judea Pearl, *Causality*: *Models, reasoning, and inference*, 2nd ed., Cambridge University Press, Cambridge, 2009. MR2548166

[5] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf, *Elements of causal inference*: *Foundations and learning algorithms*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2017. MR3822088

[6] Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn, *Algebraic statistics*: *Computational commutative algebra in statistics*, Monographs on Statistics and Applied Probability, vol. 89, Chapman & Hall/CRC, Boca Raton, FL, 2001. MR2332740

[7] Milan Studený. *Conditional independence relations have no finite complete characterization*, in S. Kubik and J.A. Visek, editors, Information Theory, Statistical Decision Functions and Random Processes. Transactions of the 11th Prague Conference, volume B, pages 377–396. Kluwer, Dordrecht, 1992.

[8] Sumio Watanabe, *Algebraic geometry and statistical learning theory*, Cambridge Monographs on Applied and Computational Mathematics, vol. 25, Cambridge University Press, Cambridge, 2009. MR2554932

Thomas Kahle

Otto-von-Guericke-Universität Magdeburg

Magdeburg, Germany

Email address: thomas.kahle@ovgu.de