The intersection axiom of conditional independence:

some “new” [?] results

Richard D. Gill
Mathematical Institute, University Leiden
This version: 26 March, 2019

\[(X \perp Y \mid Z) \& (X \perp Z \mid Y) \Rightarrow X \perp (Y, Z)\]

Presented at Algebraic Statistics seminar, Leiden, 27 February 2019

The best-laid plans of mice and men often go awry
No matter how carefully a project is planned, something may still go wrong with it.

The saying is adapted from a line in “To a Mouse,” by Robert Burns:

“The best laid schemes o' mice an' men / Gang aft a-gley.”
Algebraic Statistics

Seth Sullivant

North Carolina State University

E-mail address: smsulli2@ncsu.edu

2010 Mathematics Subject Classification. Primary 62-01, 14-01, 13P10, 13P15, 14M12, 14M25, 14P10, 14T05, 52B20, 60J10, 62F03, 62H17, 90C10, 92D15

Key words and phrases. algebraic statistics, graphical models, contingency tables, conditional independence, phylogenetic models, design of experiments, Gröbner bases, real algebraic geometry, exponential families, exact test, maximum likelihood degree, Markov basis, disclosure limitation, random graph models, model selection, identifiability

Abstract. Algebraic statistics uses tools from algebraic geometry, commutative algebra, combinatorics, and their computational sides to address problems in statistics and its applications. The starting point for this connection is the observation that many statistical models are semialgebraic sets. The algebra/statistics connection is now over twenty years old– this book presents the first comprehensive and introductory treatment of the subject. After background material in probability, algebra, and statistics, the book covers a range of topics in algebraic statistics including algebraic exponential families, likelihood inference, Fisher’s exact test, bounds on entries of contingency tables, design of experiments, identifiability of hidden variable models, phylogenetic models, and model selection. The book is suitable for both classroom use and independent study, as it has numerous examples, references, and over 150 exercises.
• The intersection axiom:
  \[(X \perp Y \mid Z) \& (X \perp Z \mid Y) \implies X \perp (Y, Z)\]

• “New” result:
  \[(X \perp Y \mid Z) \& (X \perp Z \mid Y) \iff X \perp (Y, Z) \mid W\]
  where \(W := f(Y) = g(Z)\) for some \(f, g\)

• In particular, we can take \(W = \text{Law}((Y, Z) \mid X)\)

• If \(f\) and \(g\) are trivial (constant) we obtain “axiom 5”

• Also “new”: Nontrivial \(f, g\) exist such that \(f(Y) = g(Z)\) a.e. iff \(A, B\) exist with probabilities strictly between 0 and 1 s.t.

  \[
  \Pr(Y \in A \& Z \in B^c) = 0 = \Pr(Y \in A^c \& Z \in B)
  \]

  Call such a joint law \textit{decomposable}
law(\(Y, Z\))
support(\(Z\))
<table>
<thead>
<tr>
<th></th>
<th>$\text{law}(Y_2)$</th>
<th>$\text{support}(Z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\cdot$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\cdot$</td>
</tr>
</tbody>
</table>
Comfort zones

- All variables have finite support (Algebraic Geometry)
  - All variables have countable support
- All variables have continuous joint probability densities (many applied statisticians)
  - All densities are strictly positive
- All distributions are non-degenerate Gaussian
- All variables take values in Polish spaces (My favourite)
Please recall

- The joint probability distribution of $X$ and $Y$ can be disintegrated into the marginal distribution of $X$ and a family of conditional distributions of $Y$ given $X = x$

- The disintegration is unique up to almost everywhere equivalence

- Conditional independence of $X$ and $Y$ given $Z$ is just ordinary independence within each of the joint laws of $X$ and $Y$ conditional on $Z = z$

- For me, $0/0 = \text{“undefined”}$ and $0 \times \text{“undefined”} = 0$

  - In other words: conditional distributions do exist even if we condition on zero probability events; they just fail to be uniquely defined.

  - The non-uniqueness is harmless
Some new notation

• I’ll denote by “law(X)” the probability distribution of X on \( \mathcal{X} \).

• In the finite, discrete case, a “law” is just a vector of real numbers, non-negative, adding to one.

• In the Polish case, the set of probability laws on a given Polish space is itself a Polish space under an appropriate metric. One can moreover take convex combinations.

• The family of conditional distributions of X given Y, \( (\text{law}(X \mid Y = y))_{y \in \mathcal{Y}} \) can be thought of as a function of \( y \in \mathcal{Y} \). The function in question is Borel measurable.

• As a function of the random variable Y, we can consider it as a random variable!!!!

• By Law(X \mid Y) I’ll denote that random variable, taking values in the space of probability laws on \( \mathcal{X} \).
Crucial lemma

\[ X \perp Y \mid \text{Law}(X \mid Y) \]
Lemma: $X \independent Y \mid \text{Law}(X \mid Y)$

Proof of lemma, discrete case

Recall, $X \independent Y \mid Z \iff p(x, y, z) = g(x, z) h(y, z)$

Thus $X \independent Y \mid L \iff$ we can factor $p(x, y, l)$ this way

Given function $p(x, y)$, pick any $x \in \mathcal{X}, y \in \mathcal{Y}, \ell \in \Delta_{|\mathcal{X}|-1}$

$$p(x, y, \ell) = p(x, y) \cdot 1\{\ell = p(\cdot, y)/p(y)\}$$

$$= \ell(x)p(y)1\{\ell = p(\cdot, y)/p(y)\}$$

$$= \text{Eval}(\ell, x) \cdot p(y)1\{\ell = p(\cdot, y)/p(y)\}$$

Proof of lemma, Polish case

Similar, but different – we don’t assume existence of joint densities!
Proof of forwards implication

- $X \perp Y | Z \Rightarrow \text{Law}(X | Y, Z) = \text{Law}(X | Z)$

- $X \perp Z | Y \Rightarrow \text{Law}(X | Y, Z) = \text{Law}(X | Y)$

- So we have $w(Y, Z) = g(Z) = f(Y) =: W$ for some functions $w, g, f$

- By our lemma, $X \perp (Y, Z) | \text{Law}(X | (Y, Z))$

- We found functions $g, f$ such that $g(Z) = f(Y)$ and, with $W := w(Y, Z) = g(Z) = f(Y), X \perp (Y, Z) | W$
Proof of reverse implication

- Suppose $X \perp (Y, Z) \mid W$ where $W = g(Z) = f(Y)$ for some functions $g, f$

- By axiom ..., $X \perp Y \mid (W, Z)$

- So $X \perp Y \mid (g(Z), Z)$

- So $X \perp Y \mid Z$

- Similarly, $X \perp Z \mid Y$
• Uses primary decomposition of toric ideals to come up with a nice parametrisation of the model “Axiom 5”

• Given: finite sets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, what is the set of all probability measures on their product satisfying Axiom 5, and with $p(y) > 0$, $p(z) > 0$, for all $y$, $z$?

• Answer: pick partitions of $\mathcal{Y}$, $\mathcal{Z}$ which are in 1-1 correspondence with one another. Call one of them “$\mathcal{W}$”. Pick a positive probability distribution on $\mathcal{W}$. Pick indecomposable probability distributions on the products of corresponding partition elements of $\mathcal{Y}$ and $\mathcal{Z}$. Pick probability distributions on $\mathcal{X}$, also corresponding to the preceding, not necessarily all different

• Now put them together: in simulation terms: generate r.v. $W = w \in \mathcal{W}$. Generate $(Y,Z)$ given $W = w$ and independently thereof generate $X$ given $W = w$. 
Polish spaces

- Exactly same construction ... just replace “partition” by a Borel measurable map onto another Polish space

- “Corresponding partitions” ... Borel measurable maps onto same Polish space
Questions

- Does algebraic geometry provide any further “statistical” insights?
- Can some of you join me to turn all these ideas into a nice joint paper?
- Could there be a category theoretical meta-theorem?
References

Sullivant, book, ch. 4, esp. section 4.3.1

Mathias Drton, Bernd Sturmfels, and Seth Sullivant,


References (cont.)


- van der Vaart & Wellner (1996), *Weak Convergence and Empirical Processes*

- Ghosal & van der Vaart (2017), *Fundamentals of Nonparametric Bayesian Inference*

- Aad van der Vaart (2019), personal communication
The (semi-)graphoid axioms of (conditional) independence

- **Symmetry**  \( X \perp Y \implies Y \perp X \)
- **Decomposition**  \( X \perp (Y, Z) \implies X \perp Y \)
- **Weak union**  \( X \perp (Y, Z) \implies X \perp Y \mid Z \)
- **Contraction**  \( (X \perp Z \mid Y \& X \perp Y) \implies X \perp (Y, Z) \)
- **Intersection**  \( (X \perp Y \mid Z \& X \perp Z \mid Y) \implies X \perp (Y, Z) \)
Comfort zones
All variables have:

- Finite outcome space [Nice for algebraic geometry]
- Countable outcome space
- Continuous joint density with respect to sigma-finite product measures [?]
- Outcome spaces are Polish ❤

Other “convenience” assumptions: Strictly positive joint density
Multivariate normal also allows algebraic geometry approach