The *intersection axiom* of conditional independence:

some “new” results

Richard D. Gill
Mathematical Institute, University Leiden

This version: 3 October, 2019

$$(X \perp\!\!\!\perp Y \mid Z) \;\&\; (X \perp\!\!\!\perp Z \mid Y) \quad \Longrightarrow \quad X \perp\!\!\!\perp (Y, Z)$$

Algebraic Statistics seminar, Leiden, 27 February 2019;
Combinatorics seminar 2019, SJTU, 2 October 2019

X is independent of Y given Z
and X is independent of Z given Y,
implies X is independent of Y and Z

The *intersection axiom* of conditional independence:

some "new" results

Intersection axiom.
Well known to be neither true nor even an axiom.

Richard D. Gill
Mathematical Institute, University Leiden
This version: 2 October, 2019

$$(X \perp\!\!\!\perp Y \mid Z) \;\&\; (X \perp\!\!\!\perp Z \mid Y) \quad\Longrightarrow\quad X \perp\!\!\!\perp (Y, Z)$$

Algebraic Statistics seminar, Leiden, 27 February 2019;
Combinatorics seminar 2019, JTSU, 2 October 2019

X is independent of Y given Z
and X is independent of Z given Y,
implies X is independent of Y and Z

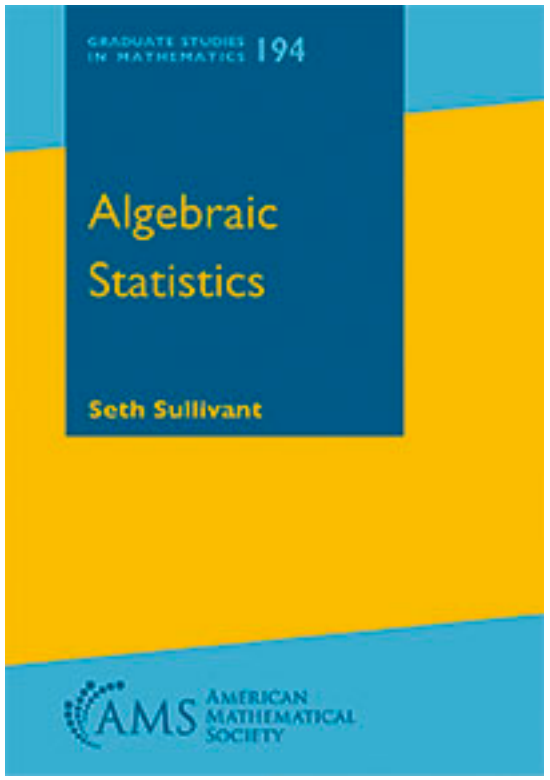
Comfort zones

All variables have:

- Finite outcome space [Nice for algebraic geometry]
- Countable outcome space
- Continuous joint density with respect to sigma-finite product measures [Usually not used rigorously]
- Outcome spaces are Polish ❤️

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

Inspiration: study group on algebraic statistics



Graduate Studies in Mathematics
Volume: 194; 2018; 490 pp; Hardcover
MSC: Primary 62; 14; 13; 52; 60; 90; 92;

Print ISBN: 978-1-4704-3517-2

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 (semi-)graphoid axioms
of (conditional) independence

1. Symmetry $X \perp\!\!\!\perp Y \implies Y \perp\!\!\!\perp X$
2. Decomposition $X \perp\!\!\!\perp (Y, Z) \implies X \perp\!\!\!\perp Y$
3. Weak union $X \perp\!\!\!\perp (Y, Z) \implies X \perp\!\!\!\perp Y \mid Z$
4. Contraction **(** $X \perp\!\!\!\perp Z \mid Y$ **&** $X \perp\!\!\!\perp Y$ **)** $\implies X \perp\!\!\!\perp (Y, Z)$
5. Intersection **(** $X \perp\!\!\!\perp Y \mid Z$ **&** $X \perp\!\!\!\perp Z \mid Y$ **)** $\implies X \perp\!\!\!\perp (Y, Z)$

1–5: (with further global conditioning): <u>the graphoid axioms</u>. Phil Dawid (1980).
1–4: (…): <u>the semi-graphoid axioms</u>
So called because of similarity to *graph separation* for subgraphs
of a simple undirected graph: A is separated from B by C

- The intersection axiom (nr 5):

$$(X \perp\!\!\!\perp Y \mid Z) \ \& \ (X \perp\!\!\!\perp Z \mid Y) \quad \Longrightarrow \quad X \perp\!\!\!\perp (Y, Z)$$

- "New" result:

$$\boxed{(X \perp\!\!\!\perp Y \mid Z) \ \& \ (X \perp\!\!\!\perp Z \mid Y) \quad \Longleftrightarrow \quad X \perp\!\!\!\perp (Y, Z) \mid W}$$

where $W := f(Y) = g(Z)$ for some f, g

- The intersection axiom:

$$(X \perp\!\!\!\perp Y \mid Z) \;\&\; (X \perp\!\!\!\perp Z \mid Y) \quad \Longrightarrow \quad X \perp\!\!\!\perp (Y, Z)$$

- "New" result:

$$\boxed{(X \perp\!\!\!\perp Y \mid Z) \;\&\; (X \perp\!\!\!\perp Z \mid Y) \quad \Longleftrightarrow \quad X \perp\!\!\!\perp (Y, Z) \mid W}$$

  where *W:= f(Y) = g(Z)* for some *f, g*

- In particular, we can take *W* = *Law((Y, Z)* | *X)*

- The intersection axiom:

$$(X \perp\!\!\!\perp Y \mid Z) \;\&\; (X \perp\!\!\!\perp Z \mid Y) \quad \Longrightarrow \quad X \perp\!\!\!\perp (Y, Z)$$

- "New" result:

$$\boxed{(X \perp\!\!\!\perp Y \mid Z) \;\&\; (X \perp\!\!\!\perp Z \mid Y) \quad \Longleftrightarrow \quad X \perp\!\!\!\perp (Y, Z) \mid W}$$

  where *W:= f(Y) = g(Z)* for some *f, g*

- In particular, we can take *W = Law((Y, Z) | X)*

- If *f* and *g* are trivial (constant) we obtain "axiom 5"

- The intersection axiom:

$$(X \perp\!\!\!\perp Y \mid Z) \ \& \ (X \perp\!\!\!\perp Z \mid Y) \quad \Longrightarrow \quad X \perp\!\!\!\perp (Y, Z)$$

- "New" result:

$$\boxed{(X \perp\!\!\!\perp Y \mid Z) \ \& \ (X \perp\!\!\!\perp Z \mid Y) \quad \Longleftrightarrow \quad X \perp\!\!\!\perp (Y, Z) \mid W}$$

  where *W:= f(Y) = g(Z)* for some *f, g*

- In particular, we can take *W = Law((Y, Z) | X)*

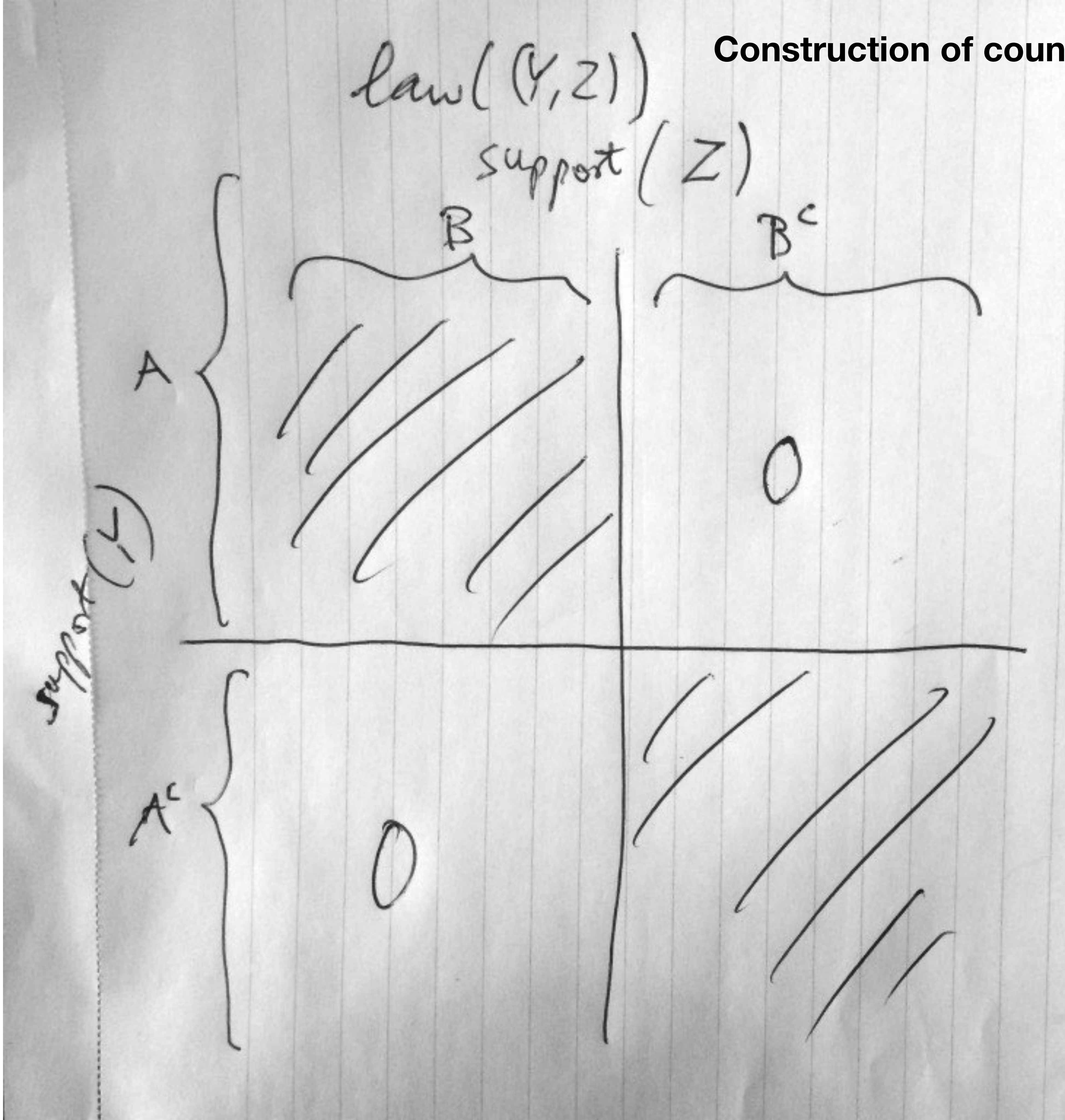
- If *f* and *g* are trivial (constant) we obtain "axiom 5"

- 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.

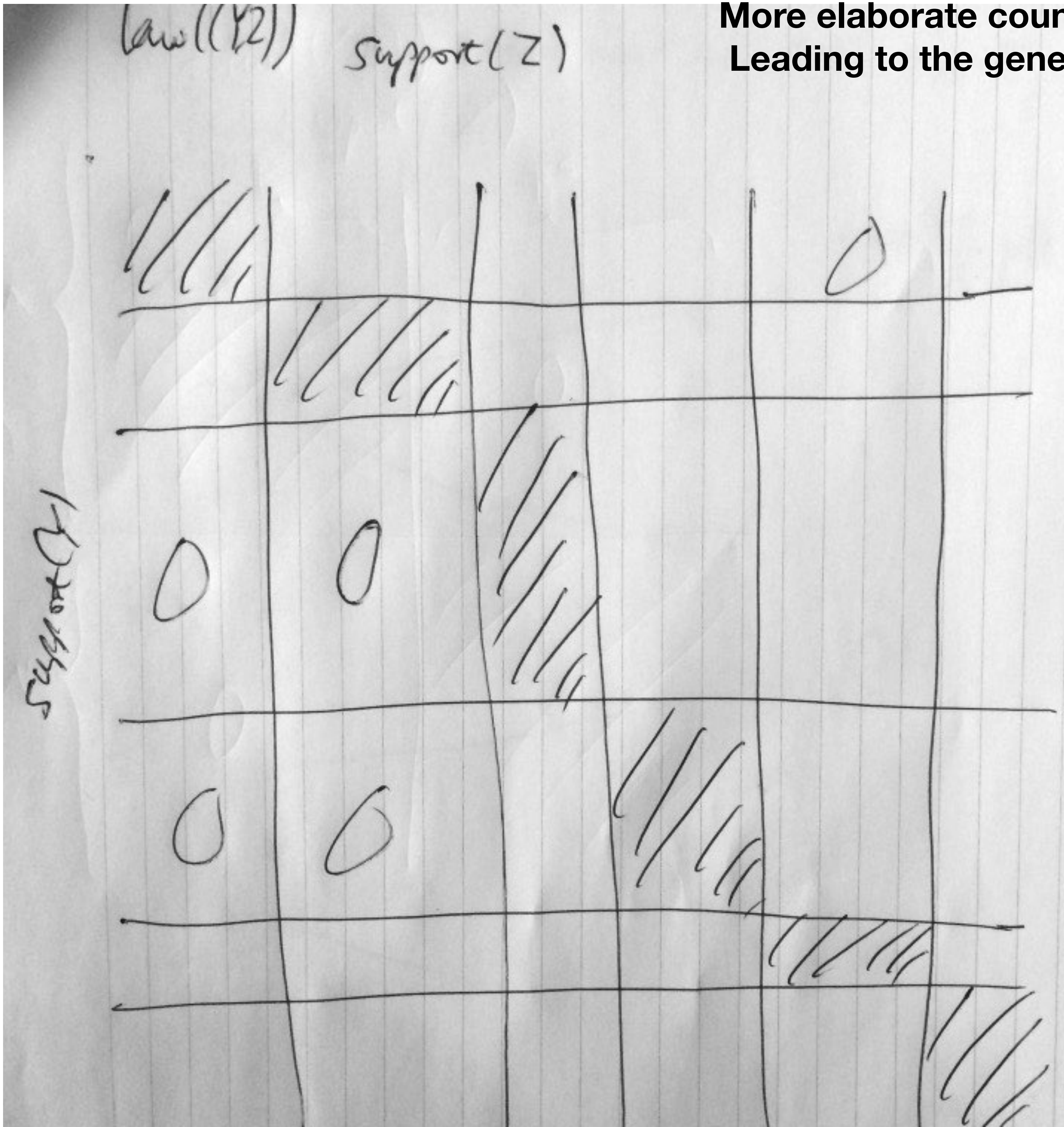
$$\Pr(Y \in A \ \& \ Z \in B^{\mathbf{c}}) \ = \ 0 \ = \ \Pr(Y \in A^{\mathbf{c}} \ \& \ Z \in B)$$

  Call such a joint law *decomposable*

Construction of counter example

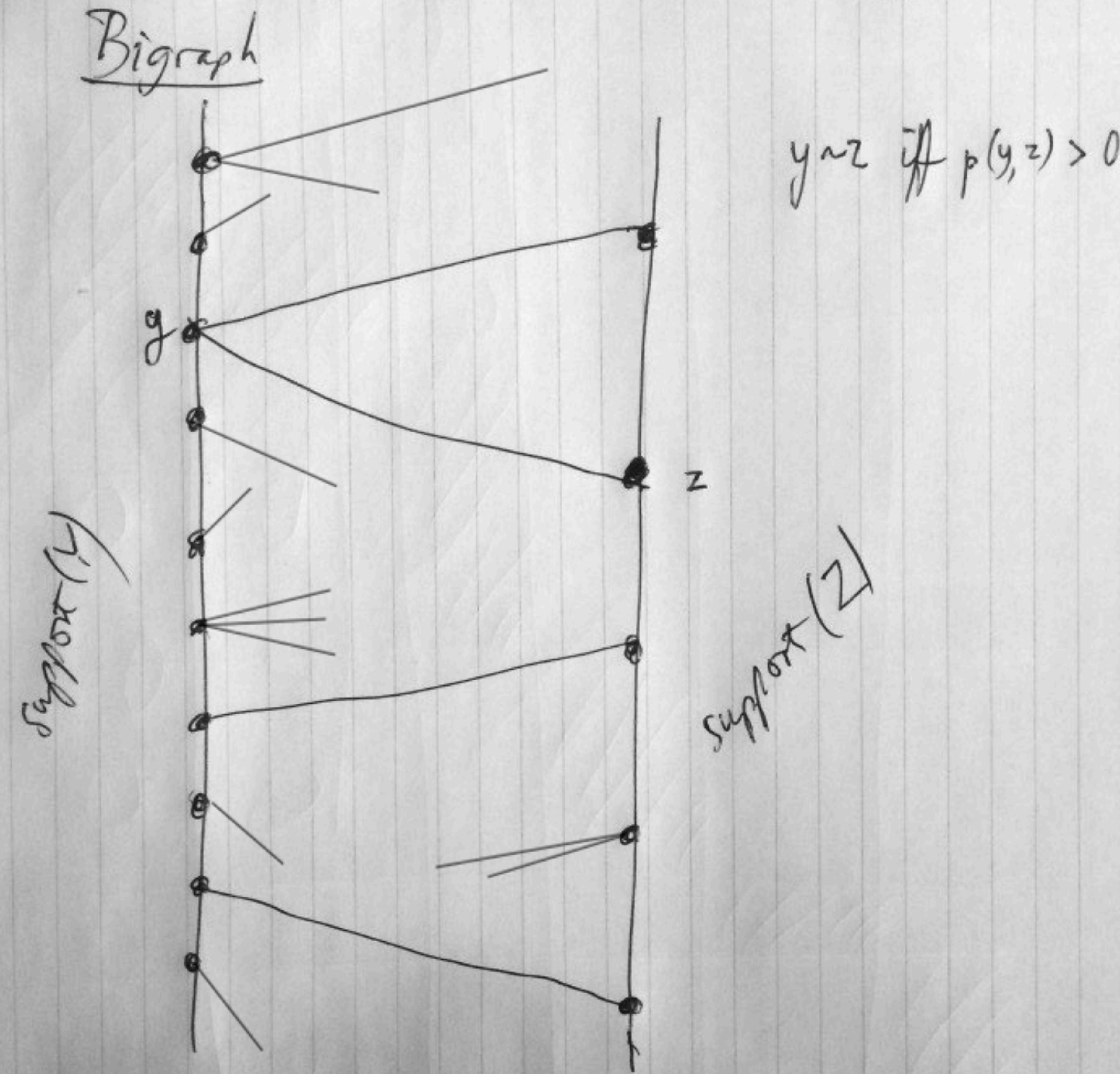


More elaborate counter example
Leading to the general theorem



Proof of new rule
Discrete case

Discrete case Y, Z each have countable support



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)

Polish space:
a topological space which can be given a metric making it complete and separable

Please recall

- The joint probability distribution of X and Y can be **disintegrated** into the marginal distribution of X and a **family** of conditional distribution**s** 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 = “undefined” and 0 x “undefined” = 0 (probability times number)

    - So: conditional distributions **do exist** if we **condition on zero probability** events; they’re just not uniquely defined.

    - The non-uniqueness is harmless

Some new notation

- I'll denote by "law(X)" the probability distribution (law) of X, where X is a random variable which takes values in a space $\mathscr{X}$. So law(X) is a probability distribution on $\mathscr{X}$

- 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, e.g., the Wasserstein metric. Disintegrations exist, Everything is nice.

- The family of conditional distributions of X given Y, $(\text{law}(X \mid Y = y))_{y \in \mathscr{Y}}$ can be thought of as a function of $y \in \mathscr{Y}$. In the Polish case, the function is Borel measurable.

- As a function of the random variable Y, we can consider it as a random variable, or as a random vector talking values in an affine space.

- By Law($X \mid Y$) I'll denote that random variable, taking values in the space of probability laws on $\mathscr{X}$.

Note distinction: Law vs. law

Crucial lemma

$$X \perp\!\!\!\perp Y \mid \mathrm{Law}(X \mid Y)$$

Lemma: $X \perp\!\!\!\perp Y \mid \mathrm{Law}(X \mid Y)$

Δ_d = probability simplex, dimension d

capital L = Law(X | Y), a random probability measure

Small ℓ ("ell") is a possible realisation

Proof of lemma, discrete case

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

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

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

$$
\begin{aligned}
p(x, y, \ell) &= p(x, y) \cdot 1\{\ell = p(\,\cdot\,, y)/p(y)\} \\
&= \ell(x) p(y) 1\{\ell = p(\,\cdot\,, y)/p(y)\} \\
&= \textbf{Eval}(\ell, x)\ .\ p(y) 1\{\ell = p(\,\cdot\,, y)/p(y)\}
\end{aligned}
$$

Proof of lemma, Polish case

Similar, but a tiny bit different – we don't assume existence of joint densities!

Proof of forwards implication

- $X \perp\!\!\!\perp Y \mid Z \implies \text{Law}(X \mid Y, Z) = \text{Law}(X \mid Z)$
- $X \perp\!\!\!\perp Z \mid Y \implies \text{Law}(X \mid Y, Z) = \text{Law}(X \mid Y)$
- So we have $w(Y, Z) = g(Z) = f(Y) =: W$ for some functions w, g, f
- By our lemma, $X \perp\!\!\!\perp (Y, Z) \mid \text{Law}(X \mid (Y, Z))$
- We found functions g, f such that $g(Z) = f(Y)$ and, with $W := w(Y, Z) = g(Z) = f(Y)$, $X \perp\!\!\!\perp (Y, Z) \mid W$

Proof of reverse implication

- Suppose $X \perp\!\!\!\perp (Y, Z) \mid W$ where $W = g(Z) = f(Y)$ for some functions g, f
- By axiom 3, $X \perp\!\!\!\perp Y \mid (W, Z)$
- So $X \perp\!\!\!\perp Y \mid (g(Z), Z)$
- So $X \perp\!\!\!\perp Y \mid Z$
- Similarly, $X \perp\!\!\!\perp Z \mid Y$

Sullivant

- Uses primary decomposition of toric ideals to come up with a nice *parametrisation* of the model "Axiom 5"

- Given: finite sets $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{Z}$, 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 $\mathscr{Y}$, $\mathscr{Z}$ which are in 1-1 correspondence with one another. Call one of them "$\mathscr{W}$". Pick a positive probability distribution on $\mathscr{W}$. Pick *indecomposable* probability distributions on the products of corresponding partition elements of $\mathscr{Y}$ and $\mathscr{Z}$. Pick probability distributions on $\mathscr{X}$, also corresponding to the preceding, not necessarily all different

- Now put them together: in simulation terms: generate r.v. $W = w \in \mathscr{W}$. 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,
______, *Lectures on Algebraic Statistics*, Oberwolfach Seminars, vol. 39, Birkhäuser Verlag, Basel, 2009. MR 2723140

Alex Fink, *The binomial ideal of the intersection axiom for conditional probabilities*, J. Algebraic Combin. **33** (2011), no. 3, 455–463. MR 2772542

Jonas Peters, *On the intersection property of conditional independence and its application to causal discovery*, Journal of Causal Inference **3** (2014), 97–108.

References (cont.)

- https://www.math.leidenuniv.nl/~vangaans/jancol1.pdf

[PDF] Probability measures on metric spaces
https://www.math.leidenuniv.nl/~vangaans/jancol1.pdf
by O van Gaans - Cited by 6 - Related articles
Onno van Gaans. These are some loose **notes** supporting the first sessions of the seminar Stochastic. Evolution ..... **space** is sometimes called a **Polish space**.

- 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