The *intersection axiom* of conditional probability:

some "new" [?] results

Richard Gill
Mathematical Institute, University Leiden
This version: 27 February, 2018

$$(X \perp Y \mid Z) \;\&\; (X \perp Z \mid Y) \quad \textbf{implies} \quad X \perp (Y, Z)$$

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."

- The intersection axiom:

$$(X \perp Y \mid Z) \ \& \ (X \perp Z \mid Y) \ \textbf{implies} \ X \perp (Y, Z)$$

- “New” result:

$$\boxed{(X \perp Y \mid Z) \ \& \ (X \perp Z \mid Y) \ \textbf{iff} \ X \perp (Y, Z) \mid W}$$

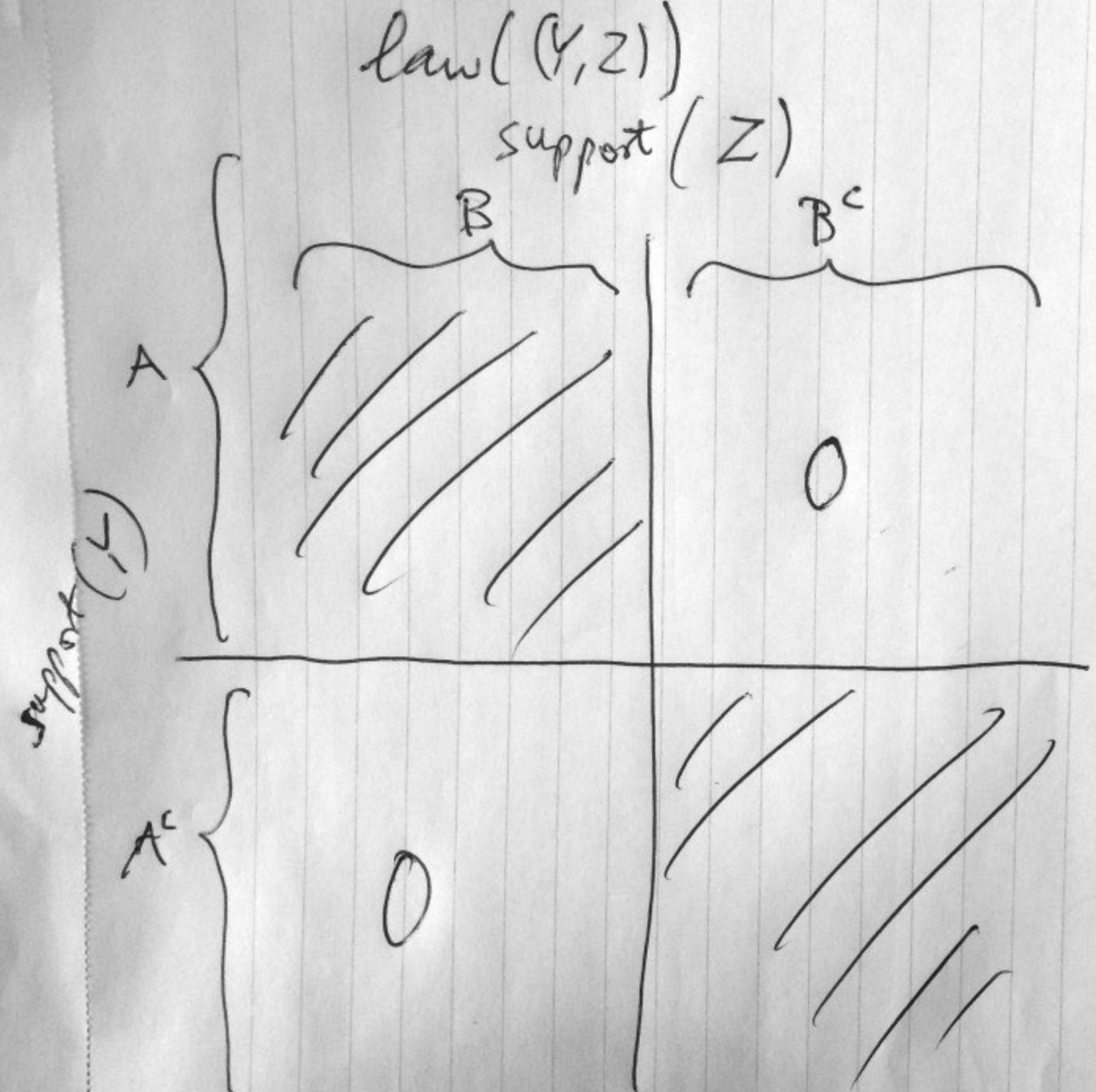
  where *W:= f(Y) = g(Z)* for some *f, g*

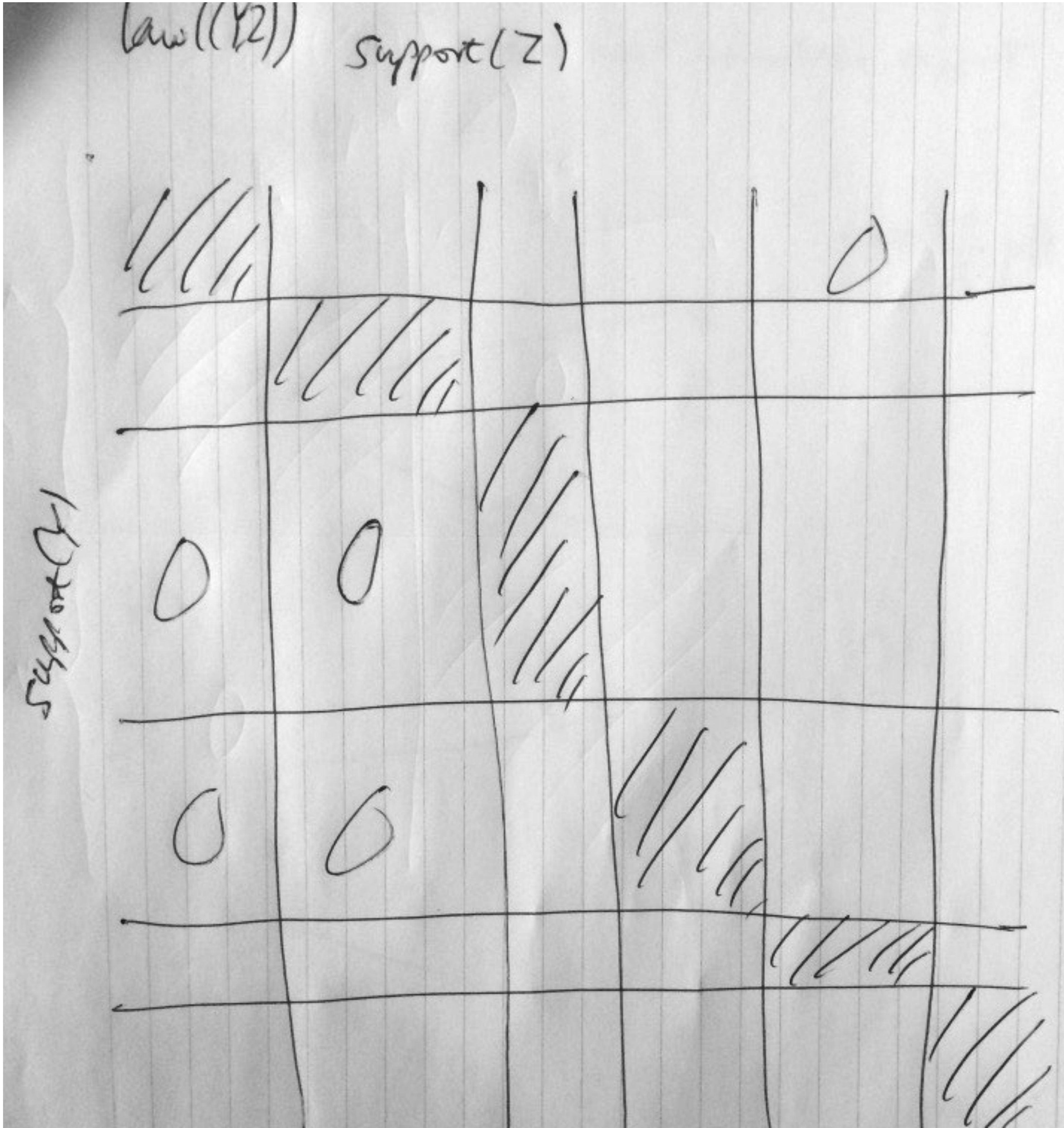
- In particular, we can take *W* = *Law((Y, Z)* | *X)*

- 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.

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

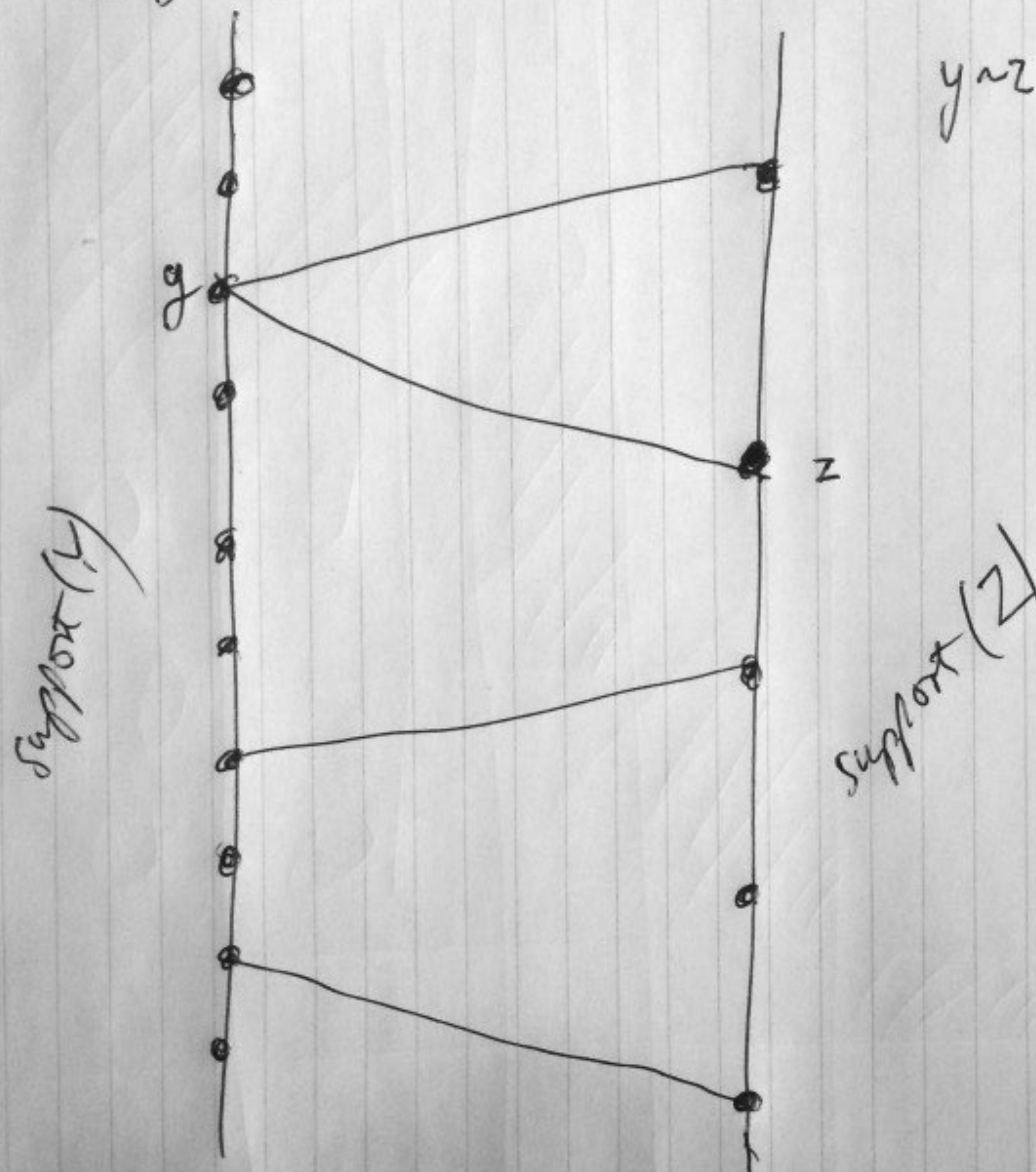

law((Y,Z))
support(Z)
B
B^c
A
A^c
support(Y)
0
0


law((YZ))
Support(Z)
Support(Y)
0
0
0
0
0

Discrete case Y, Z each have countable support

<u>Bigraph</u>

$y \sim z$ iff $p(y,z) > 0$



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 = "undefined" and 0 x "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
- In the 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, $(\text{ law}(X \mid Y = y))_y$ can be thought of as a function of 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 of X.

Crucial lemma

X is independent of Y given Law($X \mid Y$)

X is independent of Y given Law(X | Y)

Proof of lemma, discrete case

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

Thus X indep Y given L iff we can factor $p(x, y, l)$ this way

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

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

Proof of lemma, Polish case

Essentially the same proof

Proof of forwards implication

- X indep Y given Z implies Law(X | Y, Z) = Law(X | Z)
- X indep Z given Y implies Law(X | Y, Z) = Law(X | Y)
- So we have $f(Y, Z) = g(Z) = h(Y) =: W$ for some functions f, g, h
- By my lemma, X indep (Y, Z) given Law(X | (Y, Z))
- We found functions g, h such that $g(Z) = h(Y)$ and, with $W := g(Z) = h(Y)$, X indep (Y, Z) given W

Proof of reverse implication

- Suppose X indep (Y, Z) given W where $W = g(Z) = h(Y)$ for some functions g, h
- By axiom .., X indep Y given W, Z
- So X indep Y given $g(Z), Z$
- So X indep Y given Z
- Similarly, the other

Sullivant

- Uses primary decomposition of topic ideals to come up with a very nice *parametrisation* of the model “Axiom 5”

- Given: finite sets scrX, scrY, scrZ, what is the set of all probability measures on their product satisfying Axiom 5

- Answer: pick partitions of scrY, scrZ and (scrY x scrZ) which are in 1-1 correspondence with one another. Call one of them “scrW”. Pick a probability distribution on scrW. Pick *different* probability distributions on the products of corresponding partition elements of scrY and scrZ which cannot be further decomposed in this way. Pick *different* probability distributions on scrX, also corresponding to the preceding.

- Now put them together: in simulation terms: generate rv *W*. Generate (*Y,Z*) given *W* and independently thereof generate *X* given *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

~~[DSS07]~~ Mathias Drton, Bernd Sturmfels, and Seth Sullivant, *Algebraic factor analysis: tetrads, pentads and beyond*, Probab. Theory Related Fields **138** (2007), no. 3-4, 463–493.

[DSS09] ______, *Lectures on Algebraic Statistics*, Oberwolfach Seminars, vol. 39, Birkhäuser Verlag, Basel, 2009. MR 2723140

[Fin11] Alex Fink, *The binomial ideal of the intersection axiom for conditional probabilities*, J. Algebraic Combin. **33** (2011), no. 3, 455–463. MR 2772542

[Pet14] 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