Laws of conditional independence

Tobias Boege

MPI-MiS

VISTA seminar on Reusability of Research Data Across Mathematics,
University of Leipzig, 5 May 2022

Conditional independence $X \perp\!\!\!\perp Y \mid Z$

"When does knowing Z make X irrelevant for Y?"

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

- $[c_1 \perp\!\!\!\perp c_2]$
- $[c_1 \not\perp\!\!\!\perp c_2 \mid b]$ ...

Question: When can we conclude from some independences other independences?
E.g., is it possible that $[c_1 \perp\!\!\!\perp b \mid c_2]$?

Not possible!

$$[X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Z \mid Y] \Rightarrow [X \perp\!\!\!\perp Y \mid Z]$$

is a law of conditional independence (valid for all distributions).

Goal: Find all such laws. Equivalently find all patterns of conditional independence statements that can simultaneously occur ("models").

- The model of coins-and-bell is

$$[c_1 \perp\!\!\!\perp c_2] \wedge [c_1 \perp\!\!\!\perp b] \wedge [c_2 \perp\!\!\!\perp b] \wedge [c_1 \not\perp\!\!\!\perp c_2 \mid b] \wedge [c_1 \not\perp\!\!\!\perp b \mid c_2] \wedge [c_2 \not\perp\!\!\!\perp b \mid c_1].$$

Classification of models on 4 random variables

Matúš-Studený (1995–1999): 1098 models of discrete distributions.

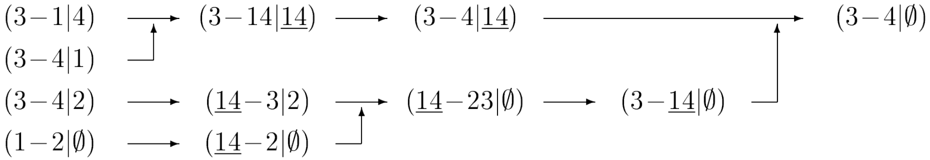
Šimeček (2006): 80 models for general Gaussian distributions.

Lněnička-Matúš (2007): 53 models for regular Gaussian distributions.

Matúš (2018): *"[...] even when N has four elements a solution is far from trivial [...] and the bin and bin$^+$-representability remain open."*

Research data: Proof schemes

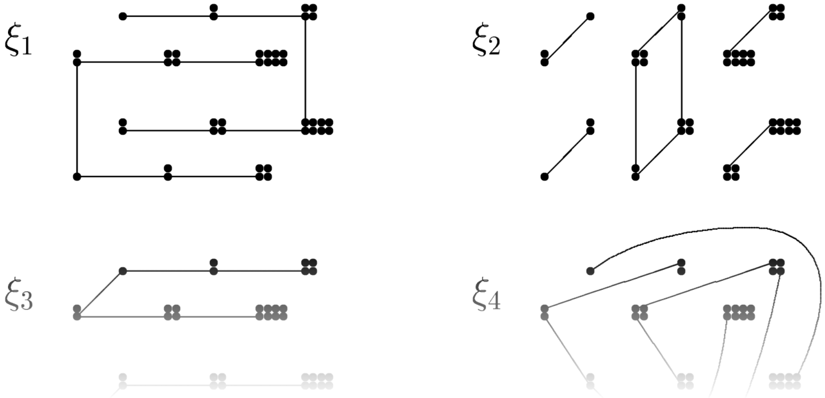
(1)



The whole proof will consist of ten schemes of the above form.

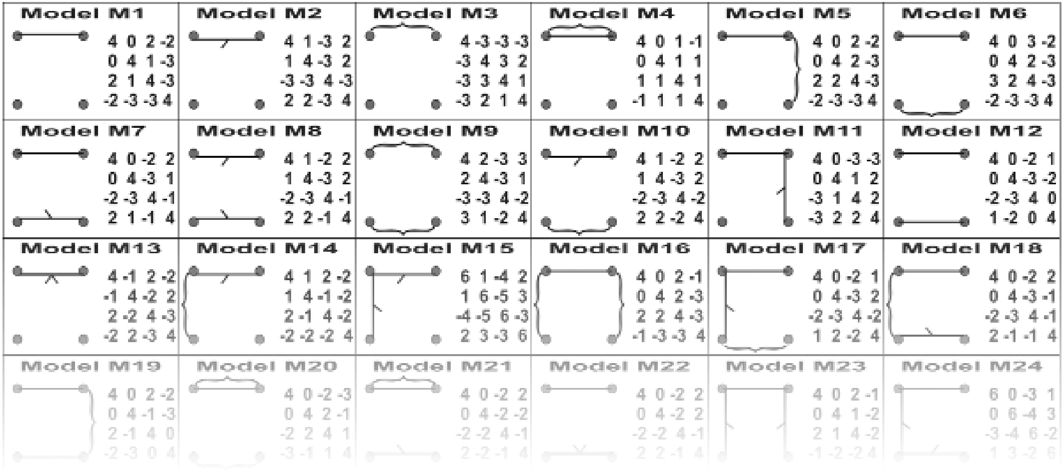
F. Matúš (1995): Conditional independences among four random variables II.

Research data: Counterexamples



F. Matúš (1995): Conditional independences among four random variables III.

Research data: Counterexamples



P. Šimeček (2006): Gaussian Representation of Independence Models over Four Random Variables.

Research data: Counterexamples

i	$A^{(i)}$	i	$A^{(i)}$	i	$A^{(i)}$	i	$A^{(i)}$
1	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon^2 \\ \varepsilon & 1 & 0 & \varepsilon \\ \varepsilon & 0 & 1 & 0 \\ \varepsilon^2 & \varepsilon & 0 & 1 \end{pmatrix}$	2	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 1 & 0 & \varepsilon^2 \\ \varepsilon & 0 & 1 & 0 \\ \varepsilon & \varepsilon^2 & 0 & 1 \end{pmatrix}$	3	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & 1\text{-}\varepsilon^2 \\ \varepsilon & 1 & 0 & \varepsilon \\ \varepsilon & 0 & 1 & 0 \\ 1\text{-}\varepsilon^2 & \varepsilon & 0 & 1 \end{pmatrix}$	4	$\begin{pmatrix} 1 & 1\text{-}\varepsilon^2 & \varepsilon^2 & 0 \\ 1\text{-}\varepsilon^2 & 1 & 0 & \varepsilon \\ \varepsilon^2 & 0 & 1 & \text{-}\varepsilon \\ 0 & \varepsilon & \text{-}\varepsilon & 1 \end{pmatrix}$
6	$\begin{pmatrix} 1 & \varepsilon^2 & \varepsilon^2 & 0 \\ \varepsilon^2 & 1 & 0 & \varepsilon \\ \varepsilon^2 & 0 & 1 & \text{-}\varepsilon \\ 0 & \varepsilon & \text{-}\varepsilon & 1 \end{pmatrix}$	7	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & 0 \\ \varepsilon & 1 & 0 & \varepsilon \\ \varepsilon & 0 & 1 & \text{-}\varepsilon \\ 0 & \varepsilon & \text{-}\varepsilon & 1 \end{pmatrix}$	8	$\begin{pmatrix} 1 & \varepsilon & \varepsilon^2 & \varepsilon \\ \varepsilon & 1 & 0 & \varepsilon \\ \varepsilon^2 & 0 & 1 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 1 \end{pmatrix}$	9	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon^2 \\ \varepsilon & 1 & 0 & \varepsilon^2 \\ \varepsilon & 0 & 1 & \varepsilon \\ \varepsilon^2 & \varepsilon^2 & \varepsilon & 1 \end{pmatrix}$
	$1 \quad \varepsilon \quad \varepsilon^2 \quad \varepsilon$		$1 \quad \varepsilon \quad \varepsilon^3 \quad \varepsilon^2$		$1 \quad \varepsilon \quad \varepsilon \quad \varepsilon^2$		$1 \quad \text{-}\varepsilon \quad \varepsilon \quad \varepsilon$

R. Lněnička & F. Matúš (2007): On Gaussian conditional independence structures.

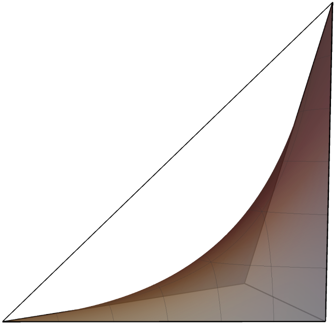
Machine-readable research data 2007 – today

http://web.archive.org/web/20070720064410/http://atrey.karlin.mff.cuni.cz/~simecek/skola/models/

- Data meanwhile deleted from institute website.
- GNU Pascal compiler for his programs hard to obtain.
- Source code comments in Czech.
- Hardly documented compiler-specific floating point data format.

→ https://github.com/taboege/simecek-tools

Conditional independence models are semialgebraic sets



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

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

$$[X \perp\!\!\!\perp Y] \Leftrightarrow p_{00} \cdot p_{11} = p_{01} \cdot p_{10},$$

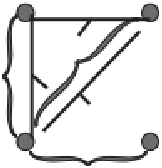
not forgetting $p_{ij} \geq 0$, $\sum p_{ij} = 1$.

Proofs and refutations on the computer

Theorem (Tarski's transfer principle)

If an inference rule is wrong, there exists a counterexample to it with real algebraic probabilities.

Model M85



1	a	b	c
a	1	d	e
b	d	1	f
c	e	f	1

Where:

$$a = \frac{3}{632836}\sqrt{1107463},$$

$$b = 10c = \frac{100}{158209}\sqrt{1107463}$$

$$d = 10e = \frac{3}{4}, f = \frac{1}{10}$$

Proofs and refutations on the computer

Theorem (Positivstellensatz)

If an inference rule is correct, there exists a proof of it in the form of a single polynomial identity with integer coefficients (called a final polynomial).

$$[X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Y \mid Z] \;\Rightarrow\; [X \perp\!\!\!\perp Z] \vee [Y \perp\!\!\!\perp Z] \;?$$

$[X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Y \mid Z] \Rightarrow [X \perp\!\!\!\perp Z] \vee [Y \perp\!\!\!\perp Z]$

$$[X \perp\!\!\!\perp Z] \cdot [Y \perp\!\!\!\perp Z] =$$

$$\big(p_{000}p_{001} + p_{001}p_{010} + p_{000}p_{011} + p_{010}p_{011} + p_{001}p_{100} + p_{011}p_{100} + p_{000}p_{101} + p_{010}p_{101} +$$

$$p_{100}p_{101} + p_{001}p_{110} + p_{011}p_{110} + p_{101}p_{110} + p_{000}p_{111} + p_{010}p_{111} + p_{100}p_{111} + p_{110}p_{111}\big) \cdot [X \perp\!\!\!\perp Y] -$$

$$\big(p_{000}p_{001} + p_{001}^2 + p_{001}p_{010} + p_{000}p_{011} + 2p_{001}p_{011} + p_{010}p_{011} + p_{011}^2 + p_{001}p_{100} + p_{011}p_{100} + p_{000}p_{101} +$$

$$2p_{001}p_{101} + p_{010}p_{101} + p_{100}p_{101} + p_{101}^2 + p_{001}p_{110} + p_{011}p_{110} + p_{101}p_{110} + p_{000}p_{111} + 4p_{001}p_{111} +$$

$$p_{010}p_{111} + 2p_{011}p_{111} + p_{100}p_{111} + 2p_{101}p_{111} + p_{110}p_{111} + p_{111}^2\big) \cdot [X \perp\!\!\!\perp Y \mid Z = 0] -$$

$$\big(p_{000}^2 + p_{000}p_{001} + 2p_{000}p_{010} + p_{001}p_{010} + p_{010}^2 + p_{000}p_{011} + p_{010}p_{011} + 2p_{000}p_{100} + p_{001}p_{100} + 4p_{010}p_{100} +$$

$$p_{011}p_{100} + p_{100}^2 + p_{000}p_{101} + p_{010}p_{101} + p_{100}p_{101} + p_{001}p_{110} + 2p_{010}p_{110} + p_{011}p_{110} + 2p_{100}p_{110} +$$

$$p_{101}p_{110} + p_{110}^2 + p_{000}p_{111} + p_{010}p_{111} + p_{100}p_{111} + p_{110}p_{111}\big) \cdot [X \perp\!\!\!\perp Y \mid Z = 1]$$

Proofs and refutations on the computer

Theorem (Tarski's transfer principle)

If an inference rule is wrong, there exists a counterexample to it with real algebraic probabilities.

Theorem (Positivstellensatz)

If an inference rule is correct, there exists a proof of it in the form of a single polynomial identity with integer coefficients (called a final polynomial).

Both are exactly representable on a computer and verifiable by off-the-shelf computer algebra systems!