Causal relations and marginal problems

Costantino Budroni

Universität Siegen

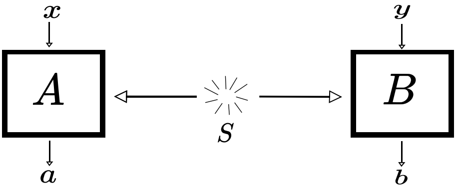
Joint work with

Nikolai Miklin (Siegen) and Rafael Chaves (Natal, Brazil)

arXiv:1607.08540

Bell theorem

Scenario[1]



Local Hidden Variable (LHV) theory

$$p(a,b|x,y) = \sum_{\lambda} p(a|x,\lambda)p(b|y,\lambda)p(\lambda)$$

Bell-CHSH inequality

$$\langle a_0 b_0 \rangle + \langle a_0 b_1 \rangle + \langle a_1 b_0 \rangle - \langle a_1 b_1 \rangle \overset{\text{LHV}}{\leq} 2$$

[1] J.S. Bell, Physics 1, (1964)

Two ingredients

Causal relations

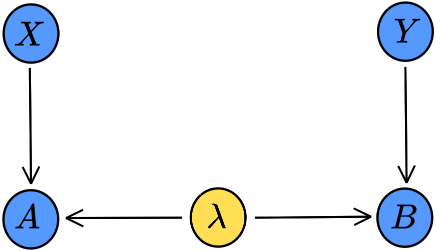
Conditional independence: $p(a, b|x, y, \lambda) = p(a|x, \lambda)p(b|y, \lambda)$
(locality)

Marginals

Limited observations: $p(a_x, b_y) := p(a, b|x, y)$
(incompatibility)

Causal structures

Graph of causal relations:



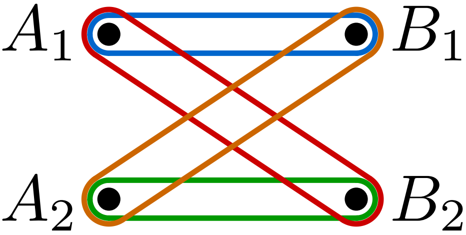
$$p(a,b,x,y,\lambda) = p(a|x,\lambda)p(b|y,\lambda)p(x)p(y)p(\lambda)$$

observed data

$$p(a,b|x,y) = \sum_{\lambda} p(a|x,\lambda)p(b|y,\lambda)p(\lambda)$$

Marginal scenario

Hypergraph of observed marginals:



$$p(a,b|x,y) = \sum_{\lambda} p(a|x,\lambda)p(b|y,\lambda)p(\lambda)$$

equivalent to (Fine's theorem[2]) marginals

$$p(a_x, b_y) = \sum_{a_{x'}, b_{y'}} p(a_x, a_{x'}, b_y, b_{y'})$$

[2] A. Fine, Phys. Rev. Lett. **48** (1982)

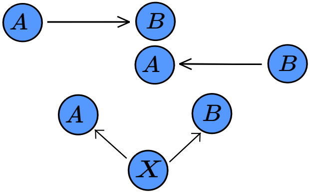
Question

What is the relation between causal structures and marginal scenarios?

Simplest example

Knowing only the marginals $p(a)$ and $p(b)$, impossible to *falsify/disprove* the causal structure (a).

(a) $p(a, b) = p(a)p(b)$



(b) $p(a, b) \neq p(a)p(b)$

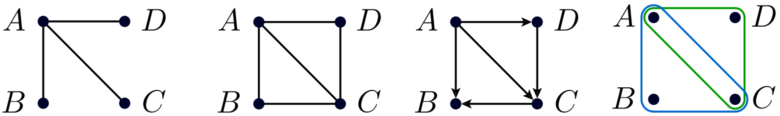
Goal

Perform this analysis for arbitrary causal structures and arbitrary marginal scenarios.

Outline of the talk

- Basic tools
  - Graphs and hypergraphs
  - Marginal scenarios
  - Causal structures (Bayesian Networks, Markov Random Fields)
- Extension of marginals (adhesivity)
- Main results on indistiguishability of causal structures
- Applications (inequalities for probability and entropy)

Graphs and hypergraphs

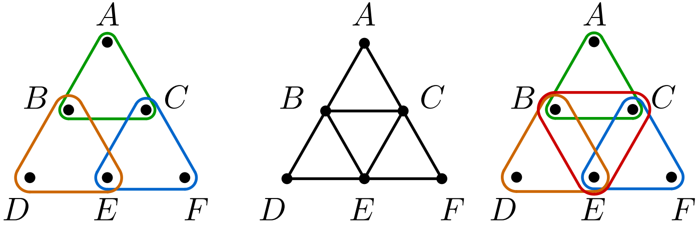


(a) $\mathcal{T}$ (b) $\mathcal{G}$ (c) DAG (d) $\mathcal{H}$

Relevant notions

- Graphs (directed/undirected).
- Hypergraphs.
- Tree graph (no cycles/loops).

Graphs and hypergraphs



(e) $\mathcal{H}$ (f) $\mathcal{G}$ (g) $\mathcal{H}'$

Relevant notions

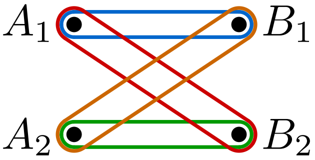
- 2-section graph (hypergraph $\rightarrow$ graph) .
- clique hypergraph (graph $\rightarrow$ hypergraph).
- Triangulated graph.
- Acyclic hypergraph $\Leftrightarrow$ clique hypergraph of a triangulated graph.
- Running intersection property (RIP): ordering of the edges $E_1, \ldots, E_n$ s. t.

$$S_i := E_i \cap (E_1 \cup \ldots \cup E_{i-1}) \subset E_j, \text{ with } j < i. \tag{1}$$

Marginal scenario

Example: CHSH scenario

Observed marginals $p(a_x, b_y)$, for $x, y = 0, 1$



$$\mathcal{M} := (\{A_1\}, \{A_2\}, \{B_1\}, \{B_2\}, \{A_1, B_1\}, \{A_1, B_2\}, \{A_2, B_1\}, \{A_2, B_2\})$$

In general

For n variables $\{X_1, \ldots, X_n\}$, $\mathcal{M} := \left\{S_1, \ldots, S_{|\mathcal{M}|}\right\}$, $S_i \subseteq \{X_1, \ldots, X_n\}$, s.t. joint probability distributions $P((X_s)_{s \in S_i})$ accessible, for each S_i

Marginal scenarios

Arise in QM from incompatibility:

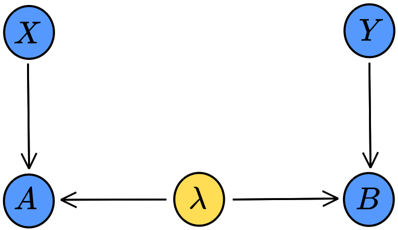
- every graph realizable as commutation-relation graph for projective measurements: connected nodes $\leftrightarrow$ commuting observables (C. Heunen, T. Fritz, M. L. Reyes, PRA **89** 2014)
- every hypergraph realizable as joint-measurability-relation graph for POVM: connected nodes $\leftrightarrow$ jointly measurable POVM (R. Kunjwal, C. Heunen and T. Fritz PRA **89** 2014)

In other fields:

- Expert systems (S. L. Lauritzen and D. J. Spiegelhalter, Journal of the Royal Statistical Society 157, 1988)
- Database theory (C. Beeri, R. Fagin, D. Maier, M. Yannakakis, Journal of the ACM 30, 1983)

Causal structures: Bayesian networks

Relations encoded in a directed acyclic graph (DAG)[3]:



Assumption:

$$p(x_1, \ldots, x_n) = \prod_i p(x_i|\mathrm{PA}_i),$$

(PA_i *parents* of i). E.g.,

$$p(a, b, x, y, \lambda) = p(a|x, \lambda)p(b|y, \lambda)p(x)p(y)p(\lambda)$$

[3] J. Pearl, Causality. Cambridge University Press, Cambridge, 2009.

Causal structures: Bayesian networks

The DAG encodes a set of conditional independence relations, e.g., Markov chain

$$(X \perp Z | Y) \;\text{ if }\; p(x, z|y) = p(x|y)p(z|y) \tag{2}$$

More generally, *d*-separation rule:

$$\mathcal{I}(\mathcal{G}) = \{(X_A \perp X_B | X_C) \mid \mathrm{dsep}_{\mathcal{G}}(A : B|C)\} \tag{3}$$

where A and B are d-separated by C, if every trail from $a \in A$ to $b \in B$, or vice versa, is *blocked* by a node in C, i.e.,

- $x \to c \to y$, or $x \leftarrow c \leftarrow y$, or $x \leftarrow c \to y$,
- or $x \to z \leftarrow y$,

for x, y, z, c in the trail, $c \in C$, and $(\{z\} \cup \mathrm{Des}_z) \cap C = \emptyset$.

---

[3] S. L. Lauritzen, *Graphical models* (Clarendon Press, 1996).

Causal structures: Markov Random Fields

Markov Random Fields (MRF) described by undirected graphs[4], e.g.,

Global Markov property

$$(X_A \perp X_B | X_C) \tag{4}$$

if every path from a node in A to a node in B passes through a node in C, denoted as $\mathrm{sep}_{\mathcal{G}}(A : B|C)$.

$$\mathcal{I}(\mathcal{G}) = \{(X_A \perp X_B | X_C) \mid \mathrm{sep}_{\mathcal{G}}(A : B|C)\}$$

[4] S. L. Lauritzen, *Graphical models* (Clarendon Press, 1996).

Extension of marginals

Main idea

Marginals may be compatible with additional independence constraints, e.g.

- $p(a), p(b)$ always compatible with $p(a, b) := p(a)p(b)$
- $p(a, b), p(b, c)$ always compatible with $p(a, b, c) := p(a|b)p(c|b)p(b) = p(a, b)p(b, c)/p(b)$.

Extension of marginals

Adhesive extension

Given $p(x_I)$ and $p'(x_J)$ coinciding on $X_{I\cap J}$

$$P(x_{I\cup J}) = \begin{cases} 0 & \text{if } p(x_{I\cap J}) = 0, \\ \frac{p(x_I)p'(x_J)}{p(x_{I\cap J})} & \text{otherwise.} \end{cases} \quad (5)$$

Result used several times in different forms [5]

[5] A. Fine, Phys. Rev. Lett. **48** (1982), C. Budroni and G. Morchio, J. Math. Phys. **51** (2010). P. Kurzyński, R. Ramanathan, and D. Kaszlikowski, Phys. Rev. Lett. **109**, (2012). J. V. Kujala, E. N. Dzhafarov, J. Å Larsson, Phys. Rev. Lett., **115** (2015)

Vorob'ev theorem

Theorem

[Vorob'ev 1963][6] *A set of probabilities associated with an acyclic marginal scenario hypergraph* $\mathcal{M}$ *admits a global extension to a single probability distribution. Moreover, the extension can be chosen as a MRF described by the 2-section graph* $[\mathcal{M}]_2$.

[6] N. Vorob'ev, Theory of Probability and Its Applications 8, 420 (1963).

Maximal set of independence conditions

Theorem

Given a joint probability distribution P on n variables $X_1, \ldots, X_n$, and a marginal scenario $\mathcal{M}$, the marginals P_{M_i} for $M_i \in \mathcal{M}$ are consistent with a probability distribution arising from a MRF associated with the 2-section graph $[\mathcal{T}]_2$, where $\mathcal{T}$ is an acyclic hypergraph extending $\mathcal{M}$.

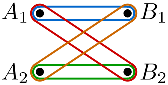
Maximal set of independence conditions

Extending hypergraph

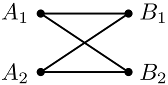
- $\mathcal{T}$ extends $\mathcal{M}$ if $\forall$ E of $\mathcal{M}$ $\exists$ E' of $\mathcal{T}$ s.t. $E \subset E'$.
- Trivial case: $\mathcal{T}$ has a single hyperedge connecting all nodes (trivial MRF).
- Computable as triangulation in linear time $O(n + e)$ for "a" minimal one.[7]
- Compute the one with the minimal number of edges (maximal number of independence constraints) is NP-hard

[7] Review: P. Heggernes, Discrete Mathematics **306**, (2006).

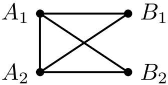
Extension of hypergraphs



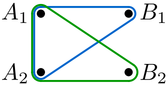
$\mathcal{M}$



$[\mathcal{M}]_2$



Triangulation



$\mathcal{T}$

CHSH example

$\{p(a_x, b_y)\}_{x,y}$ compatible with $(B_1 \perp B_2 | A_1, A_2)$, i.e.,

$$p(a_1, a_2, b_1, b_2) = p(b_1|a_1, a_2)p(b_2|a_1, a_2)p(a_1, a_2)$$

Main result

Given

- a causal structure $\mathcal{G}$, with independence relations $\mathcal{I}(\mathcal{G})$,
- a marginal scenario $\mathcal{M}$,
- his triangulations $\{\mathcal{T}_i\}_i$, and independence relations $\{\mathcal{I}(\mathcal{T}_i)\}_i$,

then

If there exits i s.t. $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(\mathcal{T}_i)$, then it is impossible to falsify $\mathcal{I}(\mathcal{G})$ given the marginals $\mathcal{M}$.

Distinction between other two cases

(a) $\forall i \ \mathcal{I}(\mathcal{T}_i) \subset \mathcal{I}(\mathcal{G})$

(b) $\forall i \ \mathcal{I}(\mathcal{G}) \not\subset \mathcal{I}(\mathcal{T}_i)$, and $\exists j$ such that $\mathcal{I}(\mathcal{T}_j) \not\subset \mathcal{I}(\mathcal{G})$.

relevant for applications, i.e., characterization of correlation polytope/entropy cone.

Applications: correlation polytopes

Unconstrained case

Vector of probabilities: $\mathbf{p} = (p_1, \ldots, p_n, \ldots, p_{ij}, \ldots, p_{ijk}, \ldots, p_{1\ldots n})$

Correlation polytope[8]: projection of the probability simplex $(p(x_1, \ldots, x_n) \geq 0, \sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) = 1)$ onto variables in $\mathcal{M}$.

- Given $\mathcal{M}$ compute a triangulation $\mathcal{T}$
- Take simplex inequalities associated with each maximal clique C_i
- Project inequalities onto $\mathcal{M}$ variables (Fourier-Motzkin elimination)

[8] I. Pitowsky, *Quantum probability – quantum logic*, (Springer-Verlag, 1989).

Applications: entropy cones

- n random variables $\{X_1, \ldots, X_n\}$
- Shannon entropy $H(X) := -\sum_x p(x) \log_2 p(x)$
- 2^n-dimensional vector $h = \left(H_\emptyset, H_{X_1}, \ldots, H_{X_i, X_j}, \ldots, H_{X_1, \ldots, X_n}\right)$.
- Entropy cone defined as:

$$\Gamma^* := \overline{\{h \in \mathbb{R}^{2^n} \mid h_S = H(S) \text{ for some entropy } H\}}.$$

- Γ^* known to be a convex cone, exact description known only for $n \leq 3$.

Applications: entropy cones

Basic inequalities for Shannon entropy

- $H(T|S) := H(S \cup T) - H(S) \geq 0$,
- $I(V : T|S) := H(V \cup S) + H(T \cup S) - H(S \cup T \cup V) - H(S) \geq 0$,

for any subset S, T, V.

Minimal set

$$\begin{aligned} H([n] \setminus \{i\}) &\leq H([n]), \\ H(S) + H(S \cup \{i,j\}) &\leq H(S \cup \{i\}) + H(S \cup \{j\}), \end{aligned} \tag{6}$$

for all $S \subset [n] \setminus \{i,j\}$, $i \neq j$ and $i,j \in [n]$, $2^{n-2}\binom{n}{2} + n$ conditions.

Shannon cone

Polyhedral cone defined by the inequality above, i.e., $\Gamma := \{h \mid Mh \geq 0\}$. Moreover, $\Gamma^* \subset \Gamma$.

---

[8] R. W. Yeung, Information theory and network coding (Springer, 2008).

Applications: entropy cones

Independence relations are nonlinear in the probabilities, e.g., Markov chain

$$p(x, z|y) = p(x|y)p(z|y)$$

- nonlinear constraints
- not a polytope, non convex
- difficult to perform variable elimination/characterize marginal scenario[9]

[9] R. Chaves Phys. Rev. Lett. **116**, 010402 (2016), D. Rosset *et al.* Phys. Rev. Lett. **116**, 010403 (2016)

Applications: entropy cones

Entropic independence relations:

$$(X \perp Z|Y) \Leftrightarrow I(X:Z|Y) = H(X,Y)+H(Z,Y)-H(X,Y,Z)-H(Y) = 0 \quad (7)$$

Linear constraint: constrained entropies still a convex cone.
Entropic characterization of a causal structure (e.g., with DAG $\mathcal{G}$)

$$\Gamma^* \cap \{I(A:B|C) = 0\}_{\mathrm{dsep}_{\mathcal{G}}(A:B|C)} \quad (8)$$

Similar approach in classical information theory: network coding[10]

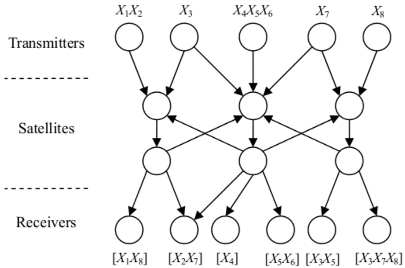


Fig. 21.7. A graph representing a satellite communication network.

[10] R. W. Yeung, Information theory and network coding (Springer, 2008).

Applications: entropy cones

Before[11]:

$$\Pi_{\mathcal{M}}\left(\Gamma^*\right) \subset \Pi_{\mathcal{M}}\left(\Gamma\right)$$

Now: given $\mathcal{M}$ and all possible triangulations $\{\mathcal{T}_i\}_i$

$$\Pi_{\mathcal{M}}\left(\Gamma^*\right) \subset \bigcap_i \Pi_{\mathcal{M}}\left(\Gamma_{\mathcal{T}_i}\right) \subset \Pi_{\mathcal{M}}\left(\Gamma\right) \subset \bigcap_i \Pi_{\mathcal{M}}\left(\bigcap_k \Gamma_{C_k^{(i)}}\right).$$

with

- $\Pi_{\mathcal{M}}$ projection on variables in $\mathcal{M}$,
- $\Gamma_{\mathcal{T}_i} := \Gamma \cap \{L_{\mathcal{I}(\mathcal{T}_i)} h = 0\}$
- $C_k^{(i)}$ maximal clique k of $\mathcal{T}_i$.
- $\Gamma_{C_k^{(i)}}$: Shannon cone for variables in $C_k^{(i)}$

[11] R. Chaves and T. Fritz, Phys. Rev. A **85** (2012); T. Fritz and R. Chaves, IEEE Trans. Inform. Theory **59**, (2013); R. Chaves, L. Luft, and D. Gross, New J. Phys. **16** (2014).

Applications: Quantum and Postquantum theories

Interesting graphs

When $\mathcal{M}$ is acyclic, classical, quantum, and postquantum marginals coincide
$\Rightarrow$ impossible to distinguish classical, quantum and postquantum theories.

Open problems

- Relation between causal structures and marginal scenarios in the quantum and postquantum case.
- Identify relevant graph theoretic notions in the quantum and postquantum case.

Conclusions

- Relation between two basic elements of Bell's theorem and its generalizations: causal structures and marginal scenarios
- Main result on indistinguishability of causal structures from limited marginals (based on the notion of adhesivity)
- Application to computation of probability and entropy inequalities for causal structures.

For more details see: arXiv:1607.08540