ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG

Information-Theoretic Approach to Causal Inference

Rafael Chaves, Lukas Luft, David Gross
Institute for Physics, University of Freiburg, Germany
http://www.qc.uni-freiburg.de

Summary

Bell's theorem in physics, as well as causal discovery in machine learning, both face the problem of deciding whether observed data is compatible with presumed causal relationships (CR). CRs can be represented by conditional independences (CI) encoded in directed acyclic graphs (DAG). The basic question can then be casted as: *given empirical data, how to decide if a presumed DAG is compatible with our observations?*

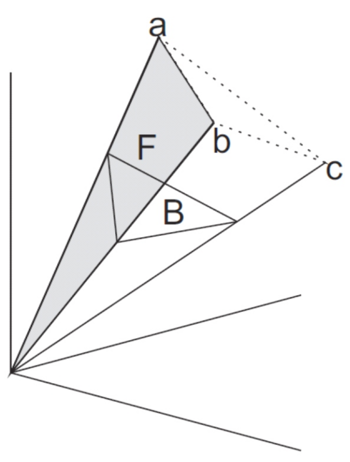
The main problem to be circumvented comes from the fact that, in terms of probabilities, CIs introduce non-linear constraints. Those lead to non-convex compatibility regions that are very difficult to be characterized.

Here, we advocate analyzing the joint entropies of observed variables for the purpose of causal inference. The entropy region associated with any given causal constraints is a convex polyhedron - a relatively simple geometric object, described completely by finitely many linear inequalities. Entropic relations naturally describe causal relationships while still retaining quantitive and useful information about causation. In this work we provide a general algorithm and discuss its application in machine learning and quantum non-locality problems.

General Framework

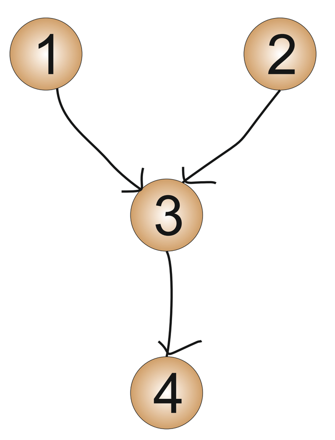
We provide a recipe that can be applied to any DAG. It consists of 3 steps:

Step 1: *Description of the unconstrained, global entropic cone*



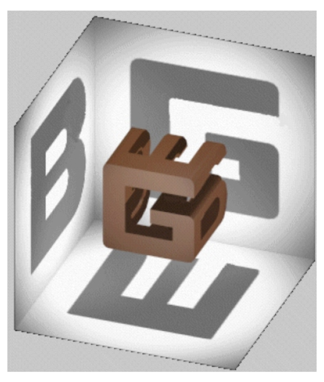
- Entropic vector $v \in \mathbb{R}^{2^n}$: each entry is the entropy $H(X_S)$ indexed by the subset $S \subset \{1,\dots,n\}$
- Defines a convex set
- Structure not fully understood, but...
- ...contained in the Shannon cone Γ_n, defined by submodularity and monotonicity

Step 2: *Choose DAG and add causal constraints*



- Conditional independences are naturally embedded in mutual informations
  $p(X_1, X_2) = p(X_1)p(X_2) \Rightarrow I(X_1, X_2) = 0$
- One can even relax (stable!)
  $I(X_1 : X_2) \leq \epsilon$
- C: cone of constraints
- $\Gamma_n \cap C$: cone of entropies subject to causal constraints

Step 3: *Marginalization*



- $\mathcal{M} \subset 2^{\{1,\dots,n\}}$: set of jointly observables
- Geometrically: just restrict $\Gamma_n \cap C$ to obs. coordinates
- Algorithmically costly: $\Gamma_n \cap C$ represented in terms of inequalities (use, e.g., Fourier-Motzkin elimination)

Final Result

- Description of marginal, causal entropic cone $(\Gamma_n \cap C)_{|\mathcal{M}}$ in terms of *„entropic Bell inequalities"*

Outlook

The information-theoretic approach

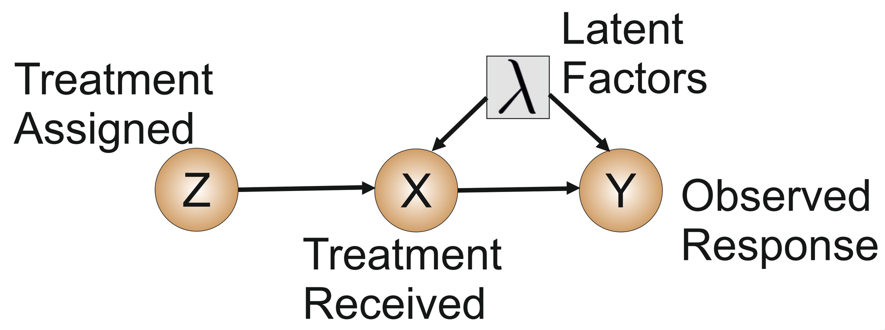
- Can be applied to derive non-trivial constraints for any graphical model
- Allows the quantification of causal influences
- Assigns an operational meaning to the violation of entropic inequalities

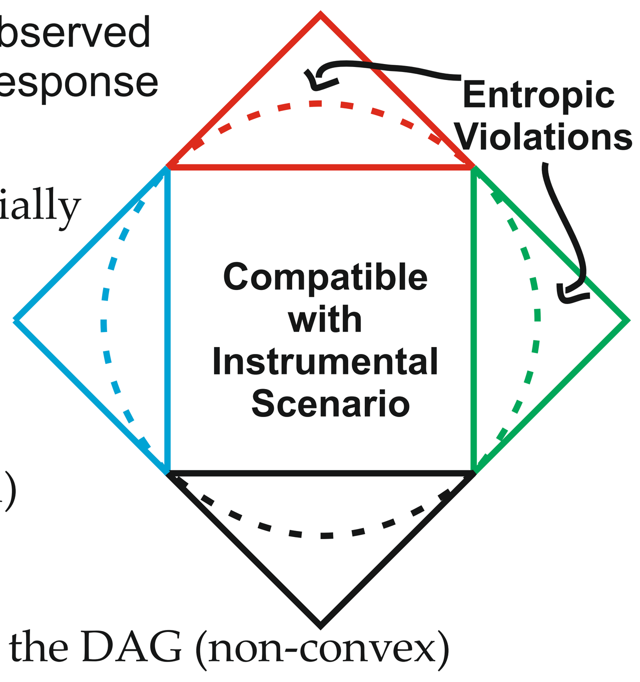
Ongoing/Future Research

- What is the complexity class of marginal problems from the entropic perspective?
- Application in time series and identifiability of homophily versus influence in social networks
- Generalization to quantum Bayesian networks. For example, Information Causality is an entropic inequality that can be derived within the framework

Results

Instrumental Inequality





- Probabilistically: number of inequalities increases exponentially
  $p(x_1, y_1|z_1) + p(x_1, y_2|z_2) \leq 1$
  $p(x_1, y_1|z_2) + p(x_1, y_2|z_1) \leq 1$
  $p(x_2, y_1|z_1) + p(x_2, y_2|z_2) \leq 1$
  $p(x_2, y_1|z_2) + p(x_2, y_2|z_1) \leq 1$
- Entropically: just one inequality (concise but coarse-grained)
  $$I(Y:Z|X) + I(X:Z) \leq H(X)$$
- Probabilistically it is very difficult to deal with variations in the DAG (non-convex)

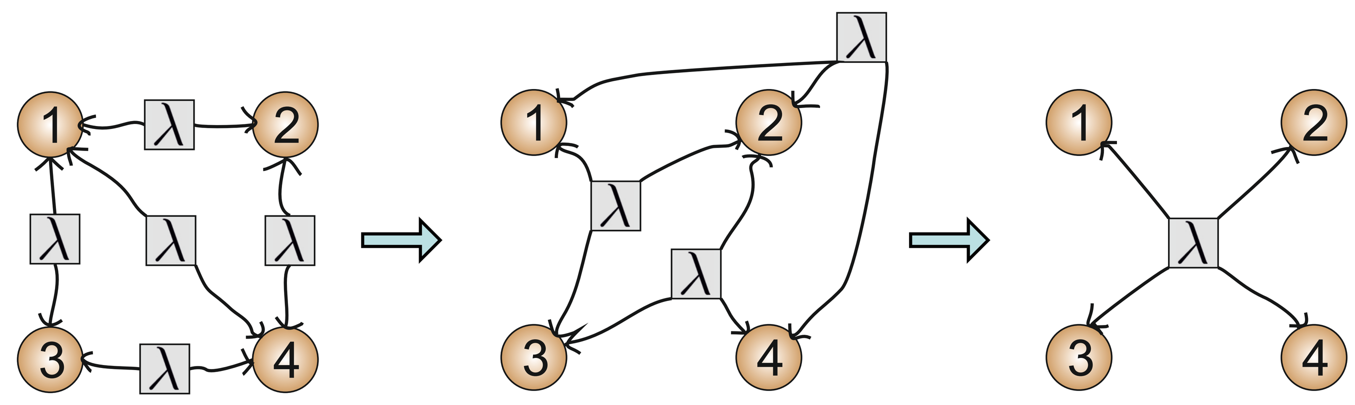
Inference of common ancestors

Can the correlations between n variables be explained by common ancestors connecting at most M of them?

$$H(\lambda_1, \lambda_2, \lambda_3) = H(\lambda_1) + H(\lambda_2) + H(\lambda_3)$$
$$H(A|\lambda_1, \lambda_2) = 0$$
$$I(A:B|\lambda_1) = 0$$
$$\mathcal{B} = I(A:B) + I(A:C) - H(A) \leq 0$$

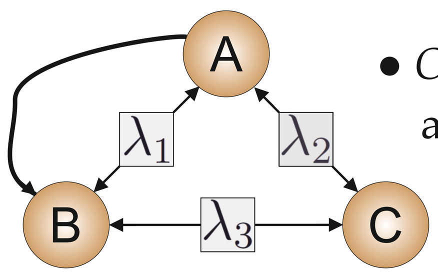
A hierarchy of causal relationship tests

$$\textstyle\sum_{i=2,\dots,n} I(X_1 : X_i) \leq (M-1)H(X_1)$$



Quantifying causal influences

- The direct causal influence $\mathcal{C}_{A\rightarrow B}$ can be lower bounded as $I(A:B|pa_B) \leq \mathcal{C}_{A\rightarrow B}$

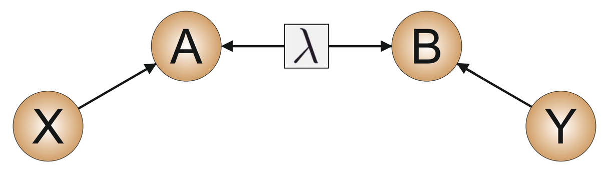


- *Operational interpretation*: violation of the entropic Bell inequality as the amount of direct influence between observable variables!
  $$\mathcal{B} \leq \mathcal{C}_{A\rightarrow B}$$

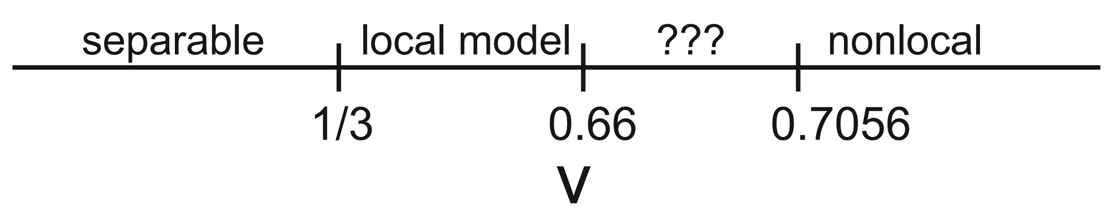
Non-locality

- Following the general framework we can derive entropic inequalities for bipartite scenarios with arbitrary setting per party (*entropic Collins-Gisin inequalities*)
- We can also derive multipartite generalizations of the Braunstein-Caves inequality
- All inequalities are valid for any number of measurement outcomes

Bell scenarios with bounded shared randomness



- Bounding the shared randomness
  $H(\lambda) \leq \mathcal{C}$
  $I(A:B|\lambda) = 0$
- „Bounded" Entropic CHSH
  $I(A_0 : B_1) + I(A_1 : B_0) - I(A_1 : B_1) - H(A_0) \leq \mathcal{C}$
- Better understanding of Werner states?
  $$\rho_W = v|\Phi^+\rangle\langle\Phi^+| + (1-v)I/4$$



References

[1] J. Pearl, *Causality* (Cambridge University Press, 2009).
[2] T. Fritz, R. Chaves, *Entropic inequalities and Marginal Problems*, IEEE Trans. Inf. Theory (2013).
[3] R. Chaves, L. Luft, D. Gross, *Causal structures from entropic information*, arxiv:1310.0284.
[4] R. Chaves, L. Luft, D. Janzing, D. Gross. *In preparation*.
[5] B. Steudel, N. Ay, *Information-theoretic inference of common ancestors, arxiv: 1010.5720.*