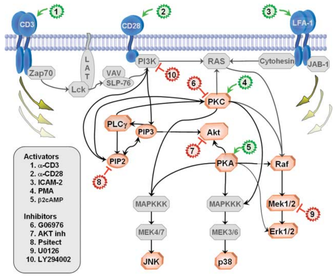
Intervention Generalization: A View from Factor Graph Models

Gecia Bravo-Hermsdorff[1], David S. Watson[2], Jialin Yu[1], Jakob Zeitler[3] & Ricardo Silva[1,3]

[1]Department of Statistical Science, University College London [2]Department of Informatics, King's College London
[3]Department of Computer Science, University College London

Motivation: extrapolating predictions to new experiments



Sach et al. (2005)

Cell biology example:
Predicting the effect of *new* combinations of gene manipulations.

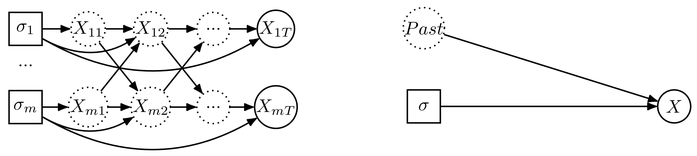
Prohibitively large combinatorial space;
unclear smoothness assumptions.

$\Longrightarrow$ Causal inference methods to the rescue!

Scope: peculiarities of our problem

Not a standard causal problem:

- no clear causal ordering (X can be undirected or contain cycles);
- interventions "shake-up" entire groups of variables, possibly overlapping;
- for each regime, data is a single snapshot of the system.



Problem set-up

Given:

- A collection of datasets collected under different regimes $\sigma \in \Sigma_{\text{train}}$.

Interventional Factor Model:

- An undirected graphical model augmented with interventional variables.
- We assume a factorization of the joint distribution that holds under all regimes:

$$p(x;\sigma) \propto \prod_{k=1}^{l} f_k(x_{S_k};\sigma_{F_k}), \qquad \forall \sigma \in \Sigma.$$

- The potential/energy functions f_k are unknown.
- The IFM describe how interventions locally changes these "soft-constraints".

Goal:

- For all *unseen* test regimes ($\sigma^\star \in \Sigma_{\text{test}} = \Sigma \backslash \Sigma_{\text{train}}$), we want to learn the density $p(x;\sigma^\star)$ (and/or predict a specific outcome).

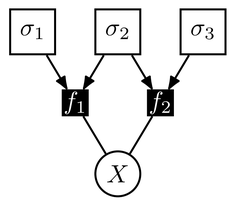
Contributions

- We introduce the interventional factor model (IFM), a novel approach for computing causal effects of *unseen* treatments (when the causal structure is messy" or otherwise uncertain).
- We establish identifiability criteria for inferring unseen treatment effects.
- Using conformal methods, we provide distribution-free predictive intervals with finite sample coverage guarantees.
- We implement efficient algorithms for learning such IFMs, and validate them on a range of semi-synthetic experiments.

Identifying a new regime (example)

Given:
this IFM model,
$p(x,\sigma) \propto f_1(x;\sigma_1,\sigma_2) f_2(x;\sigma_2,\sigma_3)$
and training data Σ_{train},
Question:
can we identify
the *unseen* regime
$\sigma^\star = (\sigma_1 = 1, \sigma_2 = 1, \sigma_3 = 1)$?

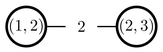
IFM:



Σ_{train}:

σ_1	σ_2	σ_3
0	0	0
0	1	0
1	0	0
1	1	0
0	0	1
0	1	1

Answer: YES!



$$\frac{p(x;(1,1,0))}{p(x;(0,1,0))} \propto \frac{f_1(x;(1,1))}{f_1(x;(0,1))} \frac{f_2(x;(1,1))}{f_2(x;(1,1))} \propto \frac{p(x;(1,1,1))}{p(x;(0,1,1))}$$

Identification results

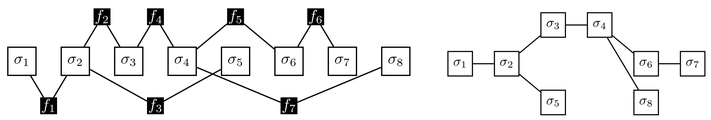
Two formulations:

1. Algebraic: finds corresponding products/ratios by solving a linear system.

	σ_1	σ_2	σ_3	f_1^{00}	f_1^{01}	f_1^{10}	f_1^{11}	f_2^{00}	f_2^{01}	f_2^{10}	f_2^{11}	f_3^{00}	f_3^{01}	f_3^{10}	f_3^{11}
q_1	0	0	0	✓				✓				✓			
q_2	0	1	0		✓					✓		✓			
q_3	1	0	0			✓		✓						✓	
q_4	1	1	0				✓			✓				✓	
q_5	0	0	1	✓					✓				✓		
q_6	0	1	1		✓				✓		✓		✓		
q_7	1	0	1			✓			✓						✓
$p^\star$	1	1	1	0	0	0	1	0	0	0	1	0	0	0	1

$$(f_1^{00} f_2^{00} f_3^{00})^{q_1} \times (f_1^{01} f_2^{10} f_3^{00})^{q_2} \times \cdots \times (f_1^{10} f_2^{01} f_3^{11})^{q_7} = f_1^{11} f_2^{11} f_3^{11}$$

2. Message-passing: if graph among intervention variables σ is decomposable.



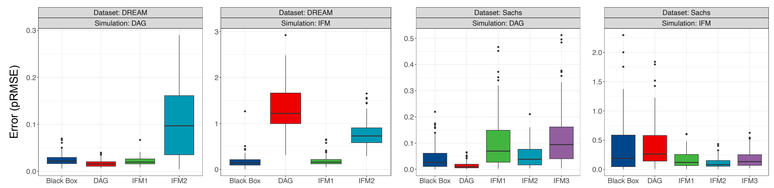
Experiments: simulations calibrated by real data

Set-up:

- Semi-synthetic data based on two biomolecular datasets: Sachs and DREAM
- For each dataset, we fit a DAG and an IFM, generating "ground-truth" outcome for each interventional regime.
  - Training data: single intervention datasets
  - Test data: combination of interventions

Compared models:

- Baseline: no structural assumptions, direct prediction from vector representation of intervention to outcome
- DAGs: correct skeleton and additive indep. noise
- IFMs: deep energy-based neural networks using: direct regression (IFM1), inverse probability weighting (IFM2), or covariate shift regression (IFM3)



Discussion

TLDR:

- IFMs offer a general approach for inferring causal effects of unseen treatments under minimal structural assumptions.

Limitations:

- DAG models fare better when the ground truth is a DAG; black box models perform well when causal effects are (approximately) linear.

Future work:

- applications to experimental design, and Bayesian optimization
- incorporating pre-treatment covariates,
- expanding to continuous interventions

Acknowledgments

We thank the anonymous reviewers and Mathias Drton for useful discussions.
GBH was supported by the ONR grant 62909-19-1-2096,
JY was supported by the EPSRC grant EP/W024330/1,
RS was partially supported by both grants,
and JZ was supported by UKRI grant EP/S021566/1.