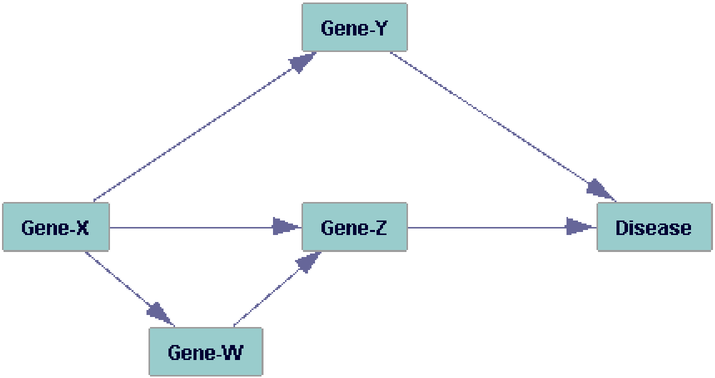
Evaluating Resampling Methods For Validating Data-Driven Causal Structures

Erich Kummerfeld, Alexander Rix

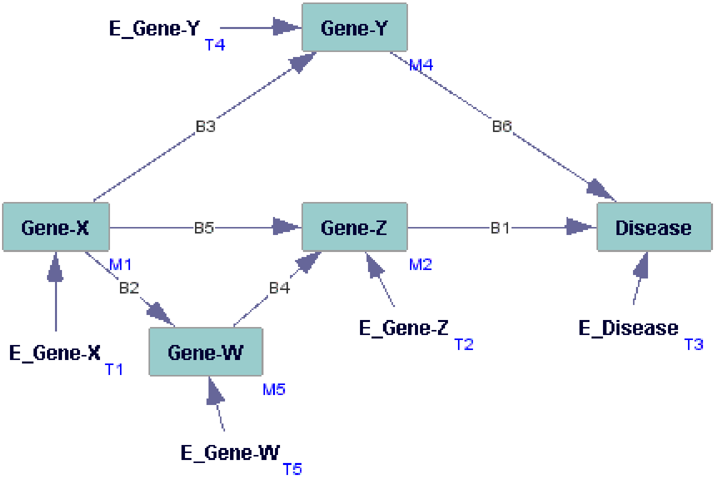
Institute for Health Informatics

May 3, 2019

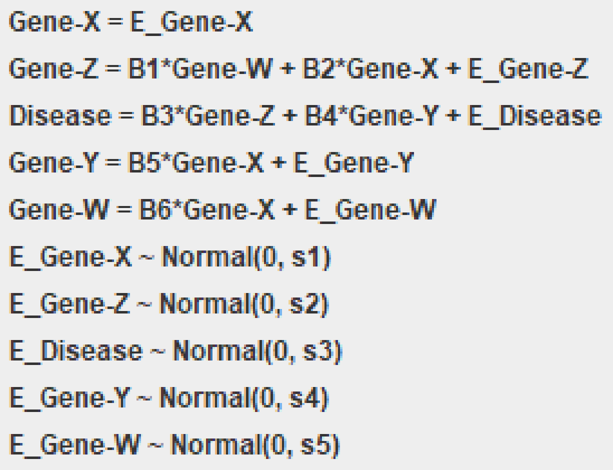
Causal Structures



Causal Parameters



Structural Equations



Gene-X = E_Gene-X

Gene-Z = B1*Gene-W + B2*Gene-X + E_Gene-Z

Disease = B3*Gene-Z + B4*Gene-Y + E_Disease

Gene-Y = B5*Gene-X + E_Gene-Y

Gene-W = B6*Gene-X + E_Gene-W

E_Gene-X ~ Normal(0, s1)

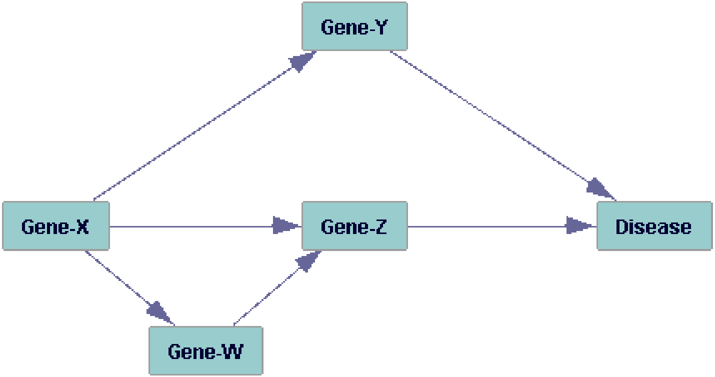
E_Gene-Z ~ Normal(0, s2)

E_Disease ~ Normal(0, s3)

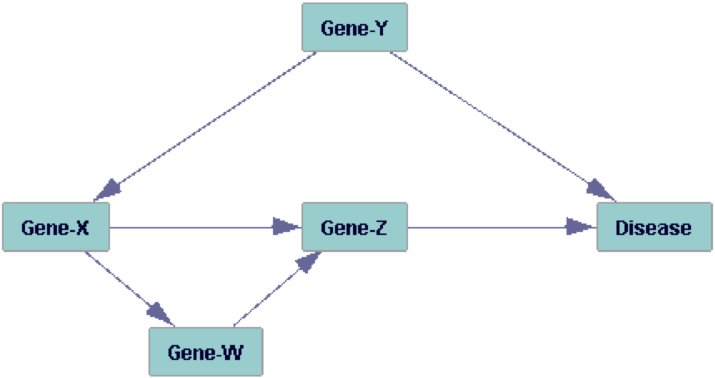
E_Gene-Y ~ Normal(0, s4)

E_Gene-W ~ Normal(0, s5)

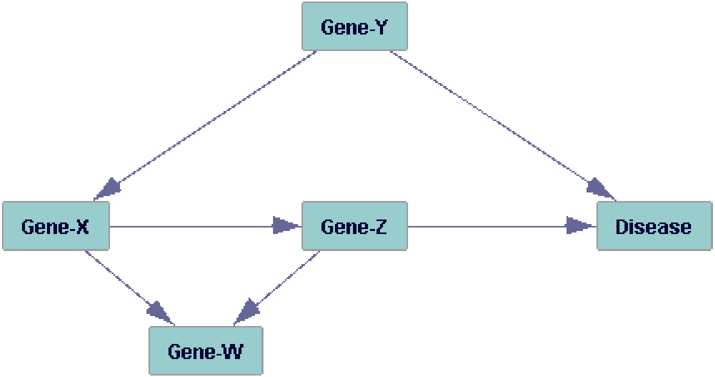
Causal Structures



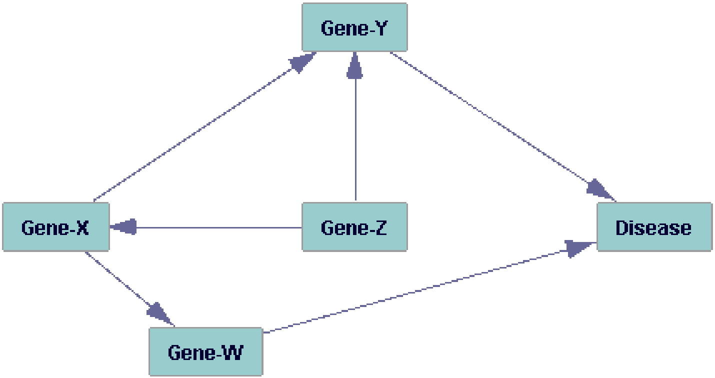
Structure Matters



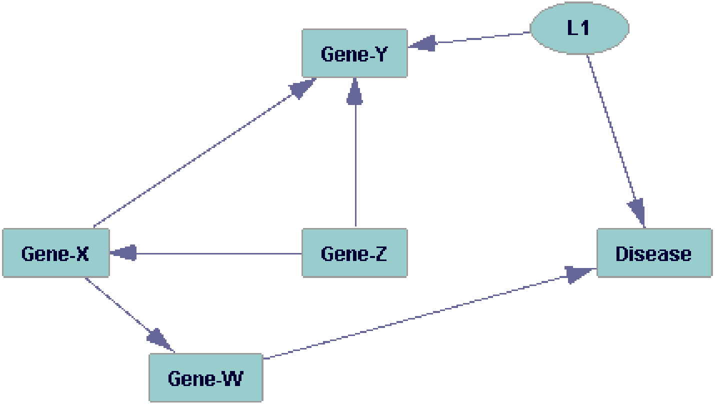
Structure Matters



Structure Matters



Structure Matters



Which Causal Structure?

What causal structure should we use for our parameter estimation?

This can be difficult: the number of DAGs scales super exponentially (roughly 2^{v^2}).

Number of DAGs over 10 variables?

Which Causal Structure?

What causal structure should we use for our parameter estimation?

This can be difficult: the number of DAGs scales super exponentially (roughly 2^{v^2}).

Number of DAGs over 10 variables?

4,175,098,976,430,598,143

Which Causal Structure?

In many cases the structure is assumed as background knowledge

- “Draw your assumptions”

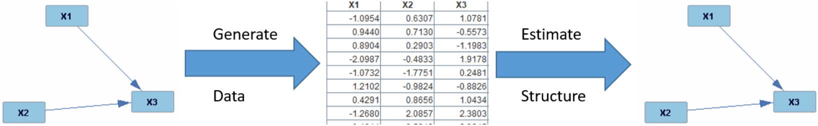
But there are also many cases where we don’t know enough to safely assume the structure.

- Many psychological data sets
- Many economic data sets
- Many genomic data sets
- Most elements of EHR data
- Most data with high numbers of variables

Causal Discovery

Causal Discovery:

- A field of study devoted to estimating, or “discovering”, an unknown causal structure from possibly observational data.
- Dozens of algorithms, with greatly varying assumptions
- Most can incorporate background knowledge
- Experimental data also allowed



X1	X2	X3
-1.0954	0.6307	1.0781
0.9440	0.7130	-0.5573
0.8904	0.2903	-1.1983
-2.0987	-0.4833	1.9178
-1.0732	-1.7751	0.2481
1.2102	-0.9824	-0.8826
0.4291	0.8656	1.0434
-1.2680	2.0857	2.3803

Causal Discovery

Not a perfect solution:

- Often identifies “equivalence class” of indistinguishable structures
- Uniform convergence proven to be impossible for any algorithm in the most general case.*
- Pointwise convergence does not give us confidence intervals.

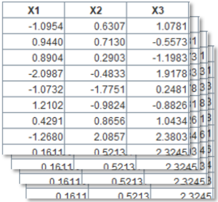
*There has been some work on circumventing this using different assumptions

Resampling

New problem: how accurate is our causal structure estimate?

One approach: resampling

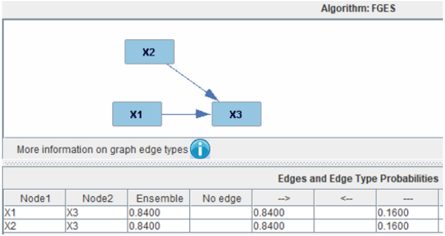
- bootstrap or jackknife the data
- how often does each possible edge appear?



X1	X2	X3
-1.0954	0.6307	1.0781
0.9440	0.7130	-0.5573
0.8904	0.2903	-1.1983
-2.0987	-0.4833	1.9178
-1.0732	-1.7751	0.2481
1.2102	-0.9824	-0.8826
0.4291	0.8656	1.0434
-1.2680	2.0857	2.3803
0.1611	0.5213	2.3245

Estimate

Structure



Algorithm: FGES

More information on graph edge types

Edges and Edge Type Probabilities

Node1	Node2	Ensemble	No edge	-->	<--	---
X1	X3	0.8400		0.8400		0.1600
X2	X3	0.8400		0.8400		0.1600

Calibration

- The frequency of times each edge occurs gives us numbers to attach to the edges.
- Higher numbers are presumably better.

Calibration

- The frequency of times each edge occurs gives us numbers to attach to the edges.
- Higher numbers are presumably better.

- But what do they mean?
- What's a good resampling frequency for an edge to have?
- If an edge appears in .7 of the resampled data sets, should we have a .7 degree of belief in that edge?
- Calibration is the correspondence between resampling frequency and justified degree of belief.

Simulation 1: Setup

Brute force approach to investigating calibration

- Randomly generated graphical models
- Randomly assign parameter values to the models
- Randomly generate data from the models

Repeat many times, and for different sample sizes.

Simulation 2: Evaluation

For each graph-data pair:

- Resample to make collection of data sets
- Estimate structure for each resampled data set
- Count proportion of times each edge occurs
- Compare to set of edges in original graph

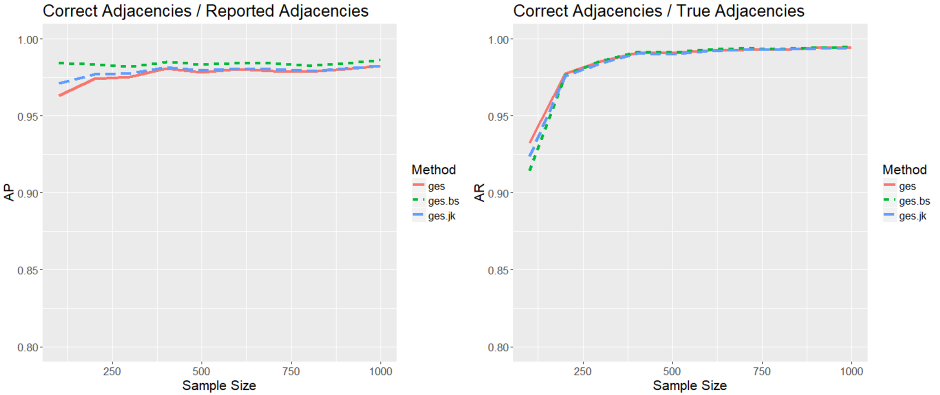
Simulation Details 1

- Data Generation Process
  - independent data sets: 500
  - number of variables: 100
  - graph generation process: 100 edges, random DAG
  - model distributional family: linear Gaussian
  - independent noise terms: Normal, 0 mean, variance drawn from Uniform(1,3)
  - edge strengths: drawn from SplitUniform(-1.5, -0.5; 0.5, 1.5)
  - sample sizes: 100, 200, ... , 900, 1000

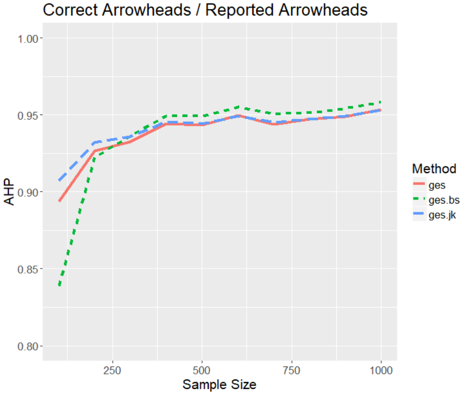
Simulation Details 2

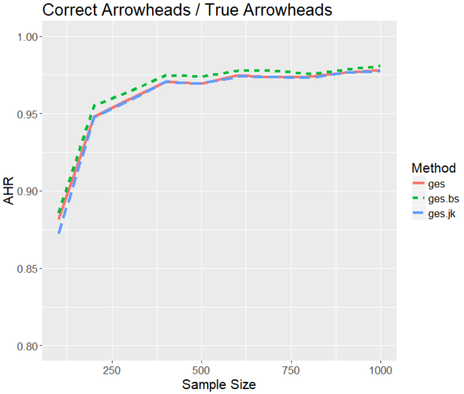
- Estimation algorithm: Greedy Equivalence Search (GES)
    - score: Bayesian Information Criterion (BIC)
    - penalty discount parameter: 2
    - other parameters: default
- Resampling
    - bootstrap: sampling with replacement, full sample size
    - jackknife: sampling without replacement, 90% sample size
    - ensemble rule: by variable pairs, highest frequency edge type

Ensemble Performance: Adjacencies

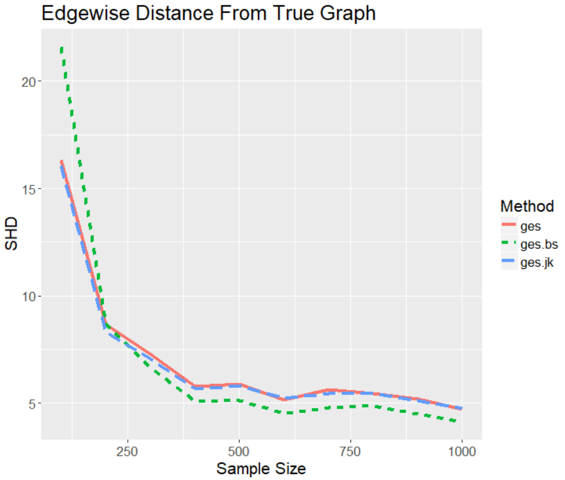


Ensemble Performance: Directionality



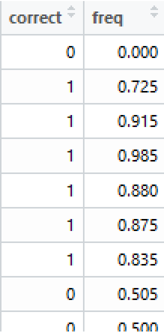


Ensemble Performance: Total Errors



Calculating Calibration 1

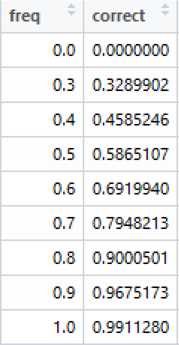
Make table of all edges from all ensemble graphs, storing their resampling frequency and whether they were correct or not.



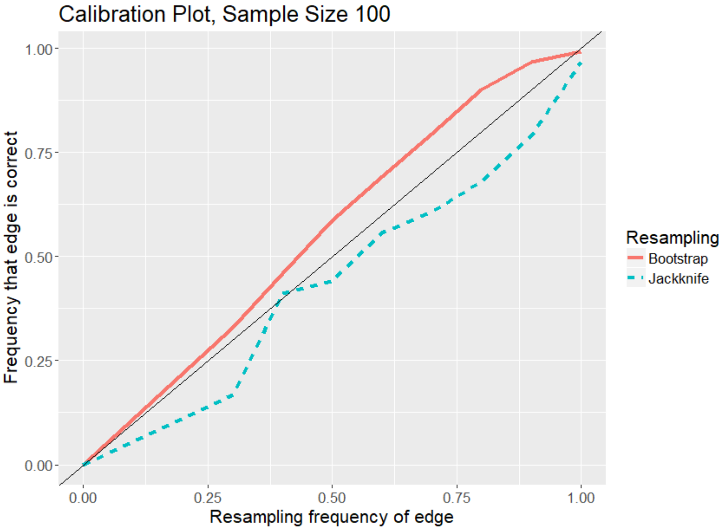
correct	freq
0	0.000
1	0.725
1	0.915
1	0.985
1	0.880
1	0.875
1	0.835
0	0.505
0	0.500

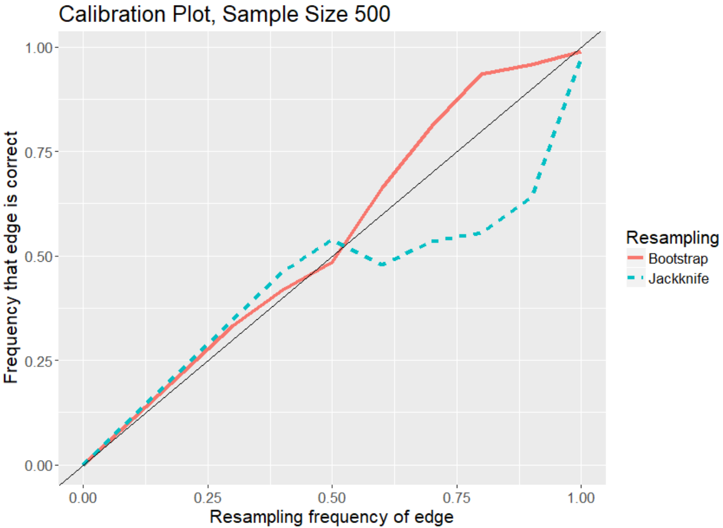
Calculating Calibration 2

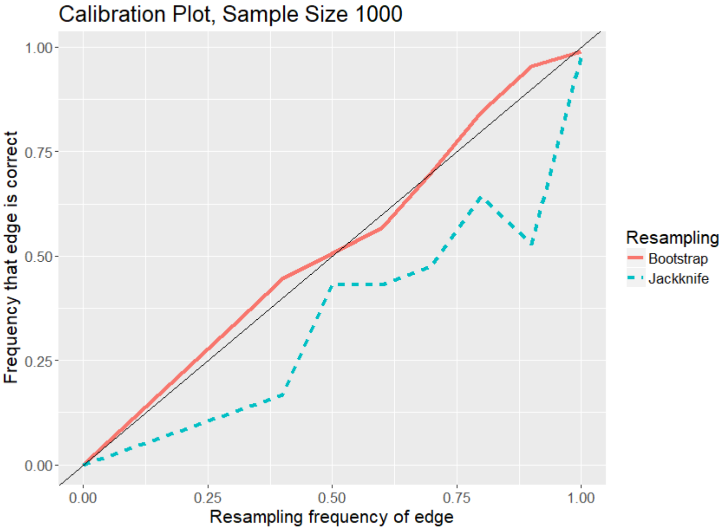
Bin the edges by their resampling frequency, calculate the proportion of edges in that bin which were correct.



freq	correct
0.0	0.0000000
0.3	0.3289902
0.4	0.4585246
0.5	0.5865107
0.6	0.6919940
0.7	0.7948213
0.8	0.9000501
0.9	0.9675173
1.0	0.9911280


Calibration Plot, Sample Size 100
Frequency that edge is correct
1.00
0.75
0.50
0.25
0.00
Resampling frequency of edge
0.00
0.25
0.50
0.75
1.00
Resampling
Bootstrap
Jackknife


Calibration Plot, Sample Size 500
Frequency that edge is correct
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Resampling frequency of edge
Resampling
Bootstrap
Jackknife


Calibration Plot, Sample Size 1000
Frequency that edge is correct
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Resampling frequency of edge
Resampling
Bootstrap
Jackknife

Limitations and Future Directions

- Other algorithms
- Other parameters
- Nonlinear relationships
- Categorical and mixed-type data
- More realistic data simulation methods
- Causal cycles
- Latent variables

Thanks!

Questions?

This project was funded in part by the University of Minnesota Clinical and Translational Science Institute.