Overview, Hard Problems, and Open Questions
Causal Inference

• What have we (hopefully) learned?
  – A representation of causal structures
  – Given a causal structure, algorithms for:
    • Predicting patterns of independence
    • Updating probabilities given observations, interventions, or both
  – Given a set of data, algorithms for:
    • Learning causal structures, potentially including:
      – Multiple unobserved common causes
      – Selection bias
      – And so on…
Causal Inference

• What else do we know how to do?
• Short answer: *Quite a lot*, including:
  – Clustering, classification, and causal learning under a range of conditions
  – Text and image classification/processing
  – Information fusion from multiple sources
  – Inference of communication networks
  – Route- or plan-generation
Causal Inference

• What do we know we cannot do?
  – Go (significantly) “beyond the data”
  – Learn when there is no variation
  – Learn efficiently in all situations
  – Reliably learn from very small samples
  – Reliably learn complex structure within an individual, given group-level data
Some Hard & Open Problems

• Can un-intervenable features be causes?
  – Not just a philosophical puzzle! Important question for policy-making decisions…
• Learning and prediction w/ definitional vars.
• Large-scale regularity and predictability from small-scale causation
• Problems of aggregation
• Distributed datasets
• Spatiotemporal data
• Feedback systems
Definitional Variables

• What happens to the Markov and Faithfulness Conditions (and thus learning and prediction) when there are definitionally related variables and (un)ambiguous manipulations?
  – Issues independently posed by Hoover and Scheines

• Short answer (Spirtes): Can sometimes learn and predict, but they are *much* harder
  – Theory is complicated and only partially worked-out
Large-scale from Small-scale

• Some seemingly innocuous observations:
  – Big stuff is made of lots of little stuff, and features of big stuff are aggregates of features of little stuff
    • There are regularities among features of little stuff
  – When we manipulate features of big stuff we are manipulating many features of little stuff
Large-scale from Small-scale

• Two questions:
  – How do the regularities among features of big stuff result from aggregation of features of little stuff and their regularities?
  – How can the regularities among features of the big stuff give us information about effects of interventions on the big stuff?
Large-scale from Small-scale

• You might be wondering: who cares?
  – We seem to be able to do prediction and
    learning with big stuff, so why worry?

• But we have principled reasons to
  question whether we really can do
  prediction and learning with big stuff…
Large-scale from Small-scale

• Take the simplest case:
  – Big-stuff is composed of lots of little-stuff
  – Every little-stuff is exactly the same
  – Every big-stuff feature is a function of the little-stuff features in the components

• The Markov properties of the big-stuff features are not (in general) the Markov properties of the little-stuff features
  – I.e., Markov properties do not necessarily hold when we aggregate lots of individuals
Large-scale from Small-scale

• Two components with the same structure

\[ X_i, Y_i, Z_i, W_i, Y_i \]

• But \( \{X_1 + X_2\} \perp \{Z_1 + Z_2\} \mid \{W_1 + W_2\}, \{Y_1 + Y_2\} \)
  - I.e., Aggregation destroys the independence
Large-scale from Small-scale

• If Markov properties change as we shift scale, then it seems that learning and prediction should only work on one level

• How can we actually do causal inference?
  – Partial answers (e.g., from Reichenbach, Strevens), but no complete theory
Distributed Datasets

• Suppose the relevant data is split into multiple datasets based on variables
  – Dataset #1 contains variables $V_1, V_2, \ldots$; Dataset #2 contains variables $W_1, W_2, \ldots$; etc.
  – Assume the variables in the datasets overlap
    • Otherwise, it isn’t a very interesting problem…

• **Central question:** What is the causal structure for *all* variables?
Distributed Datasets

• Sometimes, the datasets can potentially be unified
  – E.g., each datapoint might have a unique ID
• In these cases, there is no principled barrier to learning, since we could (if necessary) determine every datapoint’s value for every variable
  – And there has been some work designing efficient data passing algorithms
Distributed Datasets

• More interesting (and much harder) problem arises when there are no IDs
  – E.g., medical databases (from hospitals, insurers, doctors’ offices); or financial databases (from credit agencies, banks, credit card companies)

• How much information can we recover about the global causal structure?
Distributed Datasets

• Two traditional responses:
  1. Without global IDs, we can only recover structure in the local datasets, not globally
     • Overly pessimistic!
  2. Statistical Matching
     • Assume a (generic) model underlying the data, and estimate the “missing” datapoints
     • But this process assumes structural knowledge that may not be known, or may not be true
Distributed Datasets

• We can do better than these options
  – Provably correct algorithm for learning close-to-maximal information about global causal structure from multiple overlapping datasets
    • And we can sometimes learn a lot about the global structure from limited local information

  – But this algorithm is computationally hopeless
    • And its performance is not well-understood
Spatiotemporal Data

• Numerous methods for spatiotemporal data in econometrics and environmental sciences
  – But most of those methods either explicitly disavow causal interpretations, or make unreasonable assumptions

• Can we find more principled methods?

• What if the underlying generating system changes over time?
  – Can we even detect those changes reliably?
Feedback Systems

• We know how to do causal learning in feedback systems if there are no unobserved common causes
  – But not if there are such latents
• And only if we assume systems settle in equilibrium states (from which it is periodically perturbed)
  – What if that assumption is violated?