George Stepaniants

Paper Introduction

**Notes from Talk with Nathan and Bing:**

**Introduction Outline:**

Paragraph 1: (Broad Summary of Causality)

* Causality originally came from philosophy. Aristotle defined “four causes” to explanations various “Why?” questions. Other philosophers including Thomas Aquinas attributed the “first cause” of everything to God but also searched for secondary causes in natural events.
* Scientific method led us to run controlled experiments. The Neyman Rubin Causal Model was the first framework to establish how randomized experiments can be used to infer causal relationships.
* Judea Pearl gave an equivalent formulation of the Neyman-Rubin causal model through the use of Structural Equation and graphical models. Recent efforts have applied both these causal models to time-series cross-sectional data (TSCS).
* Earliest work on causal inference in time series data came from the PhD thesis of Norbert Weiner who defined that X causes Y if knowing the past of X predicts Y’s future.
* Granger later built on this notion and set it on firm footing in the context of Vector Autoregressive Models (now known as Granger Causality).
* Convergent Cross Mapping is an approach developed by George Sugihara to infer the causal relationships in complex natural systems with attractor-like structure from ecology, biology, and population sciences.
* Other methods for causal inference with time series data exist including basic correlations between time series (which can be quite effective as we see here), mutual information, Karl Friston’s Dynamical Causal Modeling for fMRI data, and graphical models such as Bayesian networks.

Paragraph 2: (Background and Methods for Time Series Network Inference)

* Networked systems are ubiquitous in practically every scientific field from economics, to psychology, biology, and physics.
* We try to represent the agents in these systems as network nodes and the directed edges that connect them signify the strength (or existence) of a causal relationship from one node to another.
* Mention important distinction between functional connectivity (correlations or similarities between time series from two nodes) and effective connectivity (determining the coupling strengths between nodes by building a good predictive physical or biological model). Granger Causality and CCM are both methods that estimate functional connectivity. However, we will show that if you have a linear system, Granger Causality is actually modeling this system and therefore is inferring its effective connectivity.
* Dynamical Causal Modeling is an inference approach that uses BOLD signals from fMRI data to reverse-engineer the connectivity of neurons in the brain. This is a great example of an inference method that determines the effective connectivity of the system.
* Another class of network inference methods perform interventions or perturbations of the network dynamics. Examples of this can be found in Marc Timme’s paper on perturbing Kuramoto networks and the large body of literature on perturbing gene regulatory networks away from equilibrium.
* In all of these inference methods, there is a great need to test them on simple systems where the ground-truth connectivity is known.
* Inference can be quite tricky and even pathological. Many problems arise with confounding factors as described in the Rubin Causal Model. Also, often the problem of network inference is not well defined as multiple networks can produce the same time dynamics (cite Barabasi’s paper on the fundamental limits of network inference).
* The Pearl and Rubin causal models are not applicable to time series network inference problems because nodes in the network are not independent. They violate the SUTVA assumption that the response of a particular unit depends only on the treatment/intervention at that one node and not on any of the other nodes around them. So we must turn to other approaches to solve this problem.
* There is a whole swath of network inference methods that simply do not work. Cite papers with figures showing blatant failure of network reconstruction.

Paragraph 3: (Granger Causality)

* Granger’s initial idea was to say that A causes B if the past of A predicts B. He formulated this rigorously through the use of Vector Autoregressive models. In Granger’s personal note on causality, he states that no one has a good definition of causality but his definition is operational.
* To apply GC to time series data from a network, it was extended to pairwise-conditional GC which essentially does n^2 comparisons to determine the significance of all possible causal connections in the network. This is the form of GC that we will focus on in this paper.
* Granger Causality has been broadly applied to the study of time series data in economics, neurobiology, genetics, etc (cite the whole list of articles that use Granger).
* Several research studies have highlighted that Granger causality does not imply real causality and does not satisfy our intuition behind causal relationships (cite papers that show evidence against GC including Bethany’s paper).

Paragraph 4: (Convergent Cross Mapping)

* Convergent Cross Mapping was made by George Sugihara to infer causal relationships in time series data using results from Take’s Delay Embedding Theorem. It is very similar to Granger Causality in its ideology and says that A causes B if history of B helps predict A.
* CCM accomplishes this by building a time-delay lagged manifold for A and a time-delay lagged manifold for B. It attempts to use points from A to predict the trajectories of points on B. If manifold A is able to predict (cross map) the trajectories on manifold B, then B encodes information about A which implies that B causes A.
* CCM has been broadly applied to problems in ecology, neurobiology, and environmental science (cite the whole swath of papers on CCM).
* There have been several studies that show that CCM is unable to infer causal structure under various regimes (cite Monster’s paper).

Paragraph 5: (Perturbation Cascade Inference and Summary of Results)

* There is a lack of literature that evaluates the network inference methods mentioned above on simple systems where the ground-truth connectivity is known. We test Granger Causality and CCM on a linear system of a network of harmonic oscillators and a nonlinear system of a network of Kuramoto oscillators.
* Granger Causality works in linear regime where data can be fit to a VAR model.
* Convergent Cross Mapping fails at the simplest linear system of harmonic oscillators. It should be used with care.
* We discuss in which regimes GC and CCM give proper inferences and why. We already know the age-old adage that “correlation is not causation”. However, it is important to understand that prediction is also not causation. This is the primary pitfall of GC and CCM. The predictive models that they fit are not the underlying models for the system and therefore, they do not learn any causal relationships.
* In this work, we show an engineering approach of how to reconstruct the network structure of a system by perturbing it and analyzing the cascades of subsequent perturbations. We call this approach Perturbation Cascade Inference (PCI)
* PCI relies on three assumptions: causes precede effects in time, are well separated in time, and we can observe effects through the use of a statistic.
* PCI works at scale for varying harmonic and Kuramoto network sizes and we analyze its performance over varying numbers of perturbations and observations.
* We also show how the performance of PCI is improved/worsened if we do an online search of nodes to perturb in the network based on data from previous perturbations.
* To conclude, if you can hit the system, then you can study perturbations and interventions of network dynamics to reverse-engineer its structure. Examples of experimentation with system perturbations occur in biology and neuroscience in the study of gene regulatory networks, drug delivery and optogenetics.

Since the birth of the scientific method, we have learned how to formulate a question and test a hypothesis through the use of an experimental study. In causal inference, the question we ask is “does X cause Y?” The earliest attempts to solve this question can be traced to the philosophical debates of … The first scientific approach to causal inference was through the use of controlled experiments. In a controlled experiment, given a parameter X and a measurement Y an intervention is performed to change X while keeping all other parameters fixed. This determines whether changing X causes a significant change in Y (i.e. if there is a causal effect). This approach has led to many fundamental discoveries in experimental physics and biology [cite lab experiments]. However, even for a controlled experiment in a laboratory setting it is often impossible to fix all the important parameters of a system.

Causal inference becomes much harder in the real world when working with large-scale studies of epidemiology, population dynamics, and social media because of the random fluctuations within samples. One of the earliest studies of causality came from the work of Jerzy Neyman who laid down the framework of potential outcomes and randomized experiments in his 1923 Master’s thesis on agricultural experimentation [cite Neyman’s thesis]. In this context, “individuals” are randomly assigned to varying levels of treatment and each individual has a different potential outcome for each treatment value. Since each individual can only be assigned one treatment, we cannot observe what the outcome for the same individual would be in a different treatment group (i.e. we cannot observe their alternative futures). This is called the “fundamental problem of causal inference” from which it naturally follows that causality cannot be deduced at the individual level. Neyman showed how the average of the individual causal effects (ICE) can be estimated in a fully randomized experiment by randomly assigning individuals to different treatment groups and taking the difference between the expected outcome in treatment and control, known as the average causal effect (ACE). This framework was then extended and formalized by Donald Rubin to deal with experiments that are not fully randomized and observational studies where interventions are not allowed. The efforts of Neyman and Rubin resulted in the Neyman-Rubin causal model, one of the earliest frameworks for determining cause and effect relationships in cross-sectional data.

In …, Judea Pearl gave an equivalent formulation of the Neyman-Rubin causal model through the use of Structural Equation and graphical models []. Recent efforts have applied both these causal models to time-series cross-sectional data (TSCS) though we do not focus on them here [].

The problem we focus on in this paper is causal inference with time series data.

Norbert Weiner first defined it in his paper by saying that X causes Y if knowing the past of X predicts Y’s future [cite Weiner’s book]. Granger further formalized this idea by phrasing it through vector autoregressive models (i.e. the simplest time-lagged linear models) [cite Granger’s personal note and original paper]. This definition of causality came to be known as “Granger Causality” and gained sufficient popularity in the applied economics community because of its clear mathematical assumptions and ease of use [cite economics papers using GC]. Many fields including healthcare, genomics, biology, and econometrics attempt to find causal relationships between components of large systems in order to understand how behavior of one component influences the behavior of the rest. Examples commonly include large neuronal circuits, gene expression networks, and the stock market. Each component of the system can naturally be represented as a node of a directed network where each directed edge from node A to node B defines a causal relationship from A to B (i.e. an increase in the price of stock A causes an increase in the price of stock B). By its original definition, Granger causality \*\*(talk about pairwise GC)\*\*. A number of research results show the success of Granger causality in inferring causal relationships in studies of such complex networked systems [cite econ and neuroscience GC papers].

Since the release of Granger’s seminal paper [cite GC here], many results have appeared showing the success of this method outside of economics in fields such as neurobiology and genetics. [cite bad GC application papers here]. However, these results rarely compare their inference

In the first part of this paper we evaluate the performance of Granger causality on physical systems where the ground-truth connectivity is known. We show that GC fails to recover the causal connections in networks of coupled harmonic and Kuramoto oscillators. We encourage the reader to study the theoretical assumptions of GC before applying it to data in physical and biological systems where those conditions may not be met.

Several research studies have highlighted that Granger causality does not imply real causality and does not satisfy our intuition behind causal relationships [cite papers that show evidence against GC including Bethany’s paper]. Granger himself stated in his personal note that “causality is a concept whose definition people know what they do not like but few know what they do like” [cite Granger’s personal note].