A Sample Path Measure of Causal Influence

Gabriel Schamberg, Todd P. Coleman

***Abstract*—**

***Index Terms*—Granger Causality, KL-Divergence, Sequential Prediction, Markov Chains.**

I. Introduction

- Granger Causality
- Directed Information
- Sequential Prediction
- Argument (neuroscience?) for why sample path causality is important

II. Sample Path Measure of Causal Influence

- Definition
- True Causal Measure is random
- Works in an online fashion

We begin by defining arbitrary measurable spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. Suppose we observe the stochastic processes $X^n \in \mathcal{X}^n$, $Y^n \in \mathcal{Y}^n$, and $Z^n \in \mathcal{Z}^n$, characterized by the joint probability density function (pdf) $f_{X^n,Y^n,Z^n}(x^n, y^n, z^n)$. We begin by considering the scenario where, having observed $(x^{i-1}, y^{i-1}, z^{i-1})$, we wish to determine the causal influence that y^{i-1} has the next observation x^i. In such a scenario, we consider the following *restricted* (denoted (r)) and *complete* (denoted (c)) conditional densities:

$$f^{(r)}_{X_i}(x_i) \triangleq f_{X_i|X^{i-1},Z^{i-1}}(x_i \mid x^{i-1}, z^{i-1}) \tag{1}$$

$$f^{(c)}_{X_i}(x_i) \triangleq f_{X_i|X^{i-1},Y^{i-1},Z^{i-1}}(x_i \mid x^{i-1}, y^{i-1}, z^{i-1}). \tag{2}$$

Using these densities, at each time i we define the sample path measure of causality from Y to X for a given realizations $(x^{i-1}, y^{i-1}, z^{i-1})$ as:

$$C_{Y\to X}(i) = D(f^{(c)}_{X_i} \mid\mid f^{(r)}_{X_i}). \tag{3}$$

The key observation that must be made that $f^{(c)}_{X_i}$ and $f^{(r)}_{X_i}$ are determined by the realizations of X, Y, and Z. As a result, *the causal measure is a random variable*. In this regard, our causal measure is different from previous measures of causality wherein the causal influence is determined by the model, and not the sample path. To ensure this point is made clear, we will present an example.

Example II.1. *Suppose $Y_i \sim$ Bern(0.5) iid for $i = 1, 2, \ldots$ and:*

$$X_i \sim \begin{cases} \textit{Bern}(0.9), & Y_{i-1} = 1 \\ \textit{Bern}(0.5), & Y_{i-1} = 0 \end{cases} \tag{4}$$

Next we note that for all $i = 1, 2, \ldots,$ we have:

$$\begin{aligned} \mathbb{P}(X_i = 1) &= \sum_{y_{i-1}\in\{0,1\}} \mathbb{P}(X_i = 1 \mid Y_{i-1} = y_{i-1})\mathbb{P}(Y_{i-1} = y_{i-1}) \\ &= (0.5)(0.5) + (0.9)(0.5) \\ &= 0.7 \end{aligned} \tag{5}$$

We can figure out the steady state distribution and the directed information rate. Then, we'll impose an initial state distribution and figure out our causal measure, which is much higher for times when $Y_{i-1} = 1$! Lastly, we can show that the information density is negative for $Y_{i-1} = 1 \to X_i = 0$.

III. Estimation of the Causal Measure

- Sequential Prediction
- Notion of Casaulity Regret
- Theorem

IV. Application to Discrete Markov Chains

- Application to conditional bernoulli model using CTW algorithm
- Explicit derivation of bounds
- Use changing environment meta algorithm for changing parameters

V. Discussion