November 07, 2019

Causality Review

The Paradigm of Kinematics and Dynamics Must Yield to Causal Structure - R. W. Spekkens (2012)

Main Thesis - The distinction of kinematics and dynamics is merely a conventional one; the pair only maintains physical significance when considered separately.

A causal account of the phenomenon provides predictions about the outcomes of interventions and the truths of counterfactuals.

Classical mechanics admits ambiguity in its choice of kinematics:
Newtonian $\longleftrightarrow$ configuration space
Hamiltonian $\longleftrightarrow$ phase space

Their empiricial reconcilation is due to an ~~comparitive~~ analogous difference in dynamical law:
Newtonian $\longleftrightarrow$ second-order Euler-Lagrange
Hamiltonian $\longleftrightarrow$ first-order Hamilton's equations.

Any modification to the dynamics of Quantum theory must remain linear (otherwise there is the possibility of superluminal signalling or violations to the second law of thermodynamics), and thus by Stinespring's dilation theorem, an appropriate extension to kinematics can revert the dynamics back to unitarity.

All of these arguments follow a unifying methodolgogical principle:
Any feature of the theory that varies among the different versions is not physical.

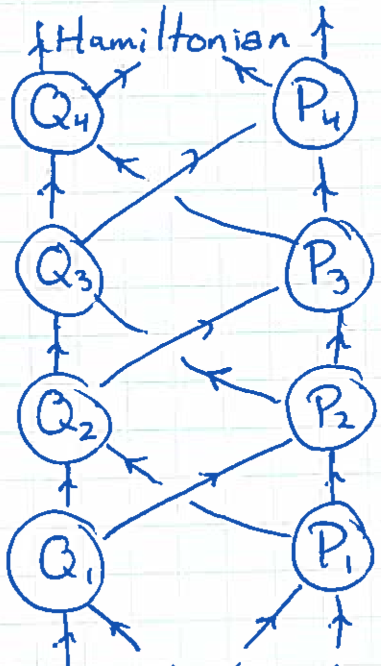
While it is possible to maintain kinematic locally by greatly violating dynamic locally, this ability is meaningless.

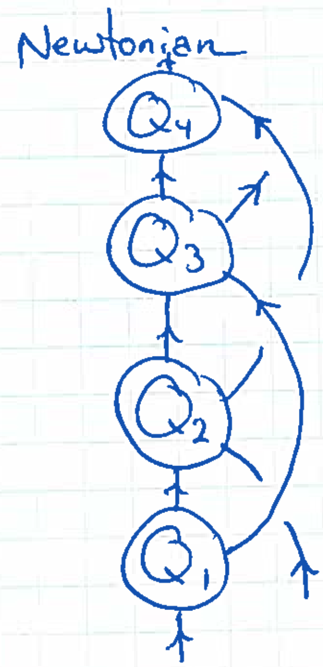
There has been significant progress in formalising mathematically the notions of causal relationships and deriving inferences from them, etc.

Primarily, statistics, machine learning, and philosophy have been carrying out these exercises.

Spirtes, Glymour, Scheines [21], and Pearl [22].

Causal Graphs for Hamiltonian and Newtonian formulations of mechanics





Here $Q_1 \not\perp Q_4 \mid Q_2, Q_3$ generically, but the specific causal-statistical parameters make this so.

Here $Q_1 \perp Q_4 \mid Q_2, Q_3$ by d-sep.

November 07, 2019

Causality Review

Detecting Causality in Complex Ecosystems - G. Sugihara, et. al. (2012)

Initial idea; ephemeral or "mirage" correlations are common in the simplest of non-linear systems.

For example;

$$X(t+1) = X(t)\left[r_x - r_x X(t) - \beta_{x,y} Y(t)\right]$$
$$Y(t+1) = Y(t)\left[r_y - r_y Y(t) - \beta_{y,x} X(t)\right]. \quad (1)$$

admits correlated intervals, anti-correlated intervals and independence (see Fig. 1).

The authors criticize Granger-causality as follows: Eq. (1) can be algebraically rearranged to give

$$X(t+1) = f_x\left(X(t), X(t-1)\right)$$

and thus information about Y becomes redundant and can be removed without affecting the predictability of $X(t+1)$. Thus Granger-causality predicts that Y does not cause X!

Takes on the fundamental assumption that causality is transitive.

$$X \Leftrightarrow Y \Leftrightarrow Z \quad \text{implies} \quad X \Leftrightarrow Z$$
$$X \Rightarrow Y, \; Y \Rightarrow Z \quad \text{implies} \quad X \Rightarrow Z$$

This is in disagreement with Pearl.

November 07, 2019

Causality Review

A branch of causality analysis tools grew out of a predictive causality principle proposed by Norbert Wiener in "The Theory of Prediction" (1956)

"Some Recent Developments In a Concept of Causality"
– C.W. J. Granger (1988) (Uses olderideas from 1963, 1980, 1987)

Considers two sets of present and future data:

$$J_t = \{(x_{t-j}, y_{t-j}, \omega_{t-j}) | j \geq 0\}$$

$$J_t' = \{(x_{t-j}, \omega_{t-j}) | j \geq 0\}$$

Where one is interested in does "y_t cause another x_t in the context of ω_t?"

Assumes that there is not function $g(\ )$ such that
$$y_t = g(\omega_{t-j}, j \geq 0)$$

$f(x | J)$ Conditional distribution of x, given J.

$E(x | J)$ conditional mean of x, given J.

Proposals (w.r.t. J_t)

(i) If $f(x_{t+1} | J_t) = f(x_{t+1} | J_t')$ then y_t does not cause x_{t+1}

(ii) If $f(x_{t+1} | J_t) \neq f(x_{t+1} | J_t')$ then y_t is a 'prima facie' cause of x_{t+1}

(iii) Same as (i) but with $f \mapsto E$, and is "in mean".

(iv) Same as (ii) but with $f \mapsto E$, and is "in mean".

Granger's Principles

(a) The cause occurs before the effect.

(b) The causal series (y_t) contains special information about the series being caused (x_t) that is not available in the other available series, here w_t.

Granger and Thompson (1987) show that (ii) implies that if one wishes to forecast x_{t+1} or any function $g(x_{t+1})$ using any cost function, then one is frequently better off using $\{y_{t-j}, j \geq 0\}$, and never worse-off.

Additional idea:

(ii)' y_t is a cause of x_{t+1} if J_t is all information in universe

Granger Writes

"It would be interesting to try to give a more formal Bayesian viewpoint of these ideas, incorporating the dynamics of prior beliefs as new information becomes available, but I do not feel competent to undertake such an analysis."

Criticisms of his approach and alternative approaches appear on p.201.

November 11, 2019

Causality Review

"Exact Inference of Causal Relations in Dynamical Systems"
- Zsigmond Benkő, et. al.

Introduction critizes two pre-existing notions of causality in dynamical systems. Predictive causality, in the sense of Granger, fails to accurately reproduce causal network structures when applied to coupled Kuramoto oscillators.

("Inferring Connectivity in Networked Dynamical Systems: Challenges Using Granger Causality").

Topological Causality methods initiated by G. Sugihara use CCM (Convergent cross mapping), which relies on Takens' Embedding Theorem. Works for circular causal relations, it remains a challenge to distinguish direct causal links from hidden common causes.

Figure 1A demonstrates that X_1 has a lower dimensionality and X_2 has a higher dimensional such that X_1 is contained in X_2. Therefore the "consequence" X_2 contains all the degrees of freedom of the cause.

The information about causal relations is contained in D_{X_1}, D_{X_2}, and D_J (where $J = X_1 X_2$ is the joint system).

Independence

$$X \perp Y \iff D_X + D_Y = D_J$$

Any interdependence makes the dimensions sub-additive.

All other possible relations are as follows:

unidirectional cause $X \rightarrow Y \iff D_X < D_Y = D_J$

circular case $X \leftrightarrow Y \iff D_X = D_Y = D_J$

common cause $X \curlyvee Y \iff \max(D_X, D_Y) < D_J < D_X + D_Y$

On pg. 6, "Theoretically, causality is a transitive relation."
But this is flawed thinking, as according to Pearl,

$$X \rightarrow Y \rightarrow Z \quad \text{and} \quad \begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & Y \arrow[r] & Z \end{tikzcd}$$

are observationally inequivalent.

They analyze the success of their framework on synthetic data.

1) Logistic Maps (Fig 1.)
2) Lorenz Systems (Fig 2.)
3) Hindmarsh-Rose Systems (Fig. 3)

And moreover extrapolate their framework to practical data to determine changes of inter-hemispheric connectivity during photo-stimulation, and also causal relations during epileptic seizures (Fig 5). (Fig 4)

In the conclusion, they say

"the Dimensional Causality method, which is the first unified way to detect all types of causal relations in dynamical systems".

$\{ \rightarrow, \leftrightarrow, \leftarrow, \vee, \perp \}$

but what about common cause and direct cause?

Correlation dimension and mutual information dimension have been applied to evaluate connections in dynamical systems, but fail to be causally exhaustive b/c of their symmetric nature 17-27.

Their "future directions" section is fascinating.

The supplementary information outlines a multitude of assumptions that go into their theory. To enforce these assumptions,

1) They ensure the two time series are stationary (augmented Dicker-Fuller unit root test)

2) Observational noise is dealt with (using filters)

3) Normalization of scales (using $[0,1]$-scaling, z-scores, quantile normalization, etc.).

In order to apply Taken's theorem to a time series w/ dimension d, one needs to select m, τ with

$$m \geq 2d+1$$

(for sufficiency) and τ. One should consult field expert, search for saturation, use false NN-method, etc. to determine m.

They start with large m and decrease m and select the smallest m which did not reduce the estimate for d too sharply.

To estimate a good τ, one could take the first zero of the autocorrelation, etc.

In general, a mountain of heuristics are applied to compute an estimate for good m and τ.

They produce manifolds for X, Y and $J = aX+Y$ w/ $a = \sqrt{\frac{29}{31}}$, (and advise against using $J=(X,Y)$). To estimate the local dimensions, they use an estimator with one parameter k, the size of neighborhood around a point to be estimated. (Some systems are sensitive to the parameter). They generate a bunch of dimension estimates and trim outliers.

Rényi dimension $\quad d_X = \lim_{N\to\infty} \frac{1}{\log N} H([X]_N) \qquad [X]_N = \frac{\lfloor NX \rfloor}{N}$

Pincus' Idea $\quad d_{X,r} = d_{X,1/N} = \frac{1}{\log N} H([X]_N)^{X \in \mathbb{R}^m}$