Causal Inference: prediction, explanation, and intervention

Lecture 6: Causality in time series (part I)
Samantha Kleinberg
samantha.kleinberg@stevens.edu
Contrast

• Exposure to someone with chickenpox Tuesday, illness Friday

• Exposure to someone with chickenpox in September, illness in November
Delays and perception

- Michotte: no perception of causality with delay
- Shanks, Pearson, and Dickinson (1999): increase in delay decreases causal judgment
Expectations

• Keyboard/triangle: Delays lower causal judgment but knowledge that there could be delay reduced effect (Buehner & May 2003)

• Energy efficient lightbulb: No effect from delay… but instant effect still judged as causal (Buehner & May 2004)

Figure 3. Mean causal ratings from Experiment 1.
- Time used to learn causality (but can incorrectly override covariation)
- More consistent vs variable timing
Temporal Predictability Facilitates Causal Learning

W. James Greville and Marc J. Buehner
Cardiff University

Temporal predictability refers to the regularity or consistency of events in time. When encountering repeated instances of causes and temporal intervals, where this interval is constant it follows from the cause. In contrast, interval variance investigated the extent to which temporal predictability facilitates causal learning. The authors demonstrated that (a) consistently judged as stronger than those with varied decline as a function of temporal uncertainty, and (c) training time. The results therefore clearly indicate temporal learning. The authors considered the implications of their findings, including associative learning theory, the attribution

Keywords: causality, predictability, contiguity, time,
The Influence of Delays in Real-Time Causal Learning

David A. Lagnado* and Maarten Speekenbrink

Department of Cognitive, Perceptual and Brain Sciences, University College London

Abstract: The close relation between time and causality is undisputed, but there is a paucity of research on how people use temporal information to inform their causal judgments. Experiment 1 examined the effect of delay variability on causal judgments, and whether participants were sensitive to the presence of a hastener cue that reduced the delay between cause and effect without changing the contingency. The results showed that higher causal ratings were given to cause-effect pairs with less variable delays, but that conditions with an active hastener actually reduced participants’ ratings of the causal cues. The latter finding can also be explained in terms of people’s sensitivity to variability, because an undetected hastener leads to greater variability in experienced delays. Experiment 2 followed up previous research showing that people give higher causal ratings to cause-effect pairs with shorter delays. We examined whether this finding might be due to the greater probability of intervening events rather than the length of delay per se. The results supported the former conjecture: participants’ causal ratings were influenced by the probability of intervening events in the cause-effect interval and not the mere length of delay. The findings from both experiments raise questions for current theories of causal learning.

Keywords: Causality judgments, temporal delays, contingency, delay variability, hasteners.
Time in graphical models

Exposure → Disease → Symptoms
First-order Markov process

• Edges only between i and i+1

\[ X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \ldots \rightarrow X_t \]

• Future and past are independent conditioned on present

\[ P(X_t \mid X_0\ldots X_{t-1}) = P(X_t \mid X_{t-1}) \]
Markov processes

- Path to current state doesn’t matter
- Once current state is known, information on prior states not informative

Diagram:
- Exposure
  - Disease
    - Symptoms
      - Treatment
Terminology

Markov process: stochastic process where transition probabilities dependent only on current state
Each node is a state (collection of variable assignments)
General

- First order: $t$ depends on $t-1$
- Second order: $t$ depends on $t-1$ and $t-2$
Hidden Markov Model

\[ P(Y_t \mid X_{0:t-1}) = P(Y_t \mid X_{t-1}) \]
Dynamic Bayesian networks

• “Dynamic” because they include time – not because they necessarily change over time!

• Generalization of HMM (HMMs are a subset of DBNs)

• Instead of one state variable, now we have a collection of BNs
Example

Initial

Transition from t to t+1
A DBN is defined by:

- a BN for the initial state of the system

- a set of BNs (one for each time slice) showing how a variable at time $t$ influences another at a later time $t+i$
Cycles

Doctor → Treatment → Outcome

D → T → O

D → T

O → t

O → t+1
Notes on DBNs

- Have to specify set of lags to test
- Complexity:
  - In theory set of graphs is all variables connected in all possible ways – across each timeslice
  - More compact than HMM: recall CPT/BN
Non-stationarity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>Marginal Likelihood</td>
<td>Marginal Likelihood</td>
<td>Marginal Likelihood</td>
<td>Marginal Likelihood</td>
<td>BIC</td>
</tr>
<tr>
<td>Change-points</td>
<td>node specific</td>
<td>whole network</td>
<td>node specific</td>
<td>whole network</td>
<td>node specific</td>
</tr>
<tr>
<td>Structure constant</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data format</td>
<td>Continuous</td>
<td>Discrete</td>
<td>Continuous</td>
<td>Continuous</td>
<td>Continuous</td>
</tr>
<tr>
<td>Latent variables</td>
<td>Change-point process</td>
<td>Change-point process</td>
<td>Change-point process</td>
<td>Free allocation</td>
<td>Free allocation</td>
</tr>
</tbody>
</table>

Some software

- Banjo (java) – discretized data
- DBmcmc, BNT, DBNbox (MATLAB)
- miniTUBA
- GlobalMIT, libDAI (C++)

Application: gene regulatory networks

Application: neuronal

But...

- Nonstationarity
- Data collection granularity
- Post hoc ergo propter hoc
- Missing data
Nonstationarity

The graph shows a plot of US Highway fatalities against Lemons imported. The data points suggest a trend, but the variability indicates nonstationarity in the dataset.
• Properties change over time
  • Mean, variance, etc
Potential solutions

• First, test for stationarity
  • unit root test (e.g. Dickey-Fuller)

• Find stationary subset

• Make the data stationary
  • differencing each day
  • differencing across years (for seasonal trends)
Correlation: 0.8204

Correlation: 0.4714
data not missing completely at random

- how frequently do you measure your temperature when you are not sick?
Logic-based

- Relationships versus structures
- Complexity + time
- Finding time windows without prior knowledge
- Automated explanation with uncertainty + time

The following slides are from:
What is a causal relationship?

smoking and hypertension until a particular genetic mutation occurs causes CHF in 1-2 years with probability $p$

$$(s \land h)Ug \xrightarrow{\geq 1, \leq 2} CHF$$

*made up relationship + timing
How to distinguish cause/non-cause

• Causes raise probability of effects
• But so do non-causes
Significance/insignificance

Suppes: $P(e|c^d) = P(e|c)$ so $d$ spurious
BN: Distribution unfaithful
Eells: wouldn't hold $d$ fixed when evaluating $c$
What do we want from a measure of causal significance?

• Proportional to how “informative” cause is

• Direct causes

• Distinguish between strong/weak causes
Causal significance

- Potential causes raise probability $P(e|c) > P(e)$ or change expected value $E[e|c] \neq E[e]$  
  
- Assess average difference cause makes to probability of effect

\[
\varepsilon_{avg}(c, e) = \frac{\sum_{x \in X} c P(e|c \land x) - P(e|\neg c \land x)}{|X \setminus c|}
\]

- Note time:
\( e_{\text{avg}} \) and time

\[ c \rightarrow^{{\geq s, \leq t}} e \]
\[ x \rightarrow^{{\geq p}} e \]

\[ P(e|c^x) \] means \( c, x \) both occur and \( e \) occurs in overlapping time window.
Finding timing

• Main idea: looking for better explanations for the effect

• Inferring timing

  • Instead of accepting/rejecting hypotheses, refine them from data

  • Can start by testing relationships between all variables and CHF in 1-2 weeks, and ultimately infer "high AST leads to CHF in 4-10 days"
Finding timings: greedy search

1) Too wide
2) Shifted
3) Too narrow

\[ \varepsilon_{avg}(c, e) = \frac{\sum_{x \in X} c P(e|c \land x) - P(e|\neg c \land x)}{|X \setminus c|} \]

\[ P(e|c \land x) = \frac{\#(c \land x \land e)}{\#(c \land x)} \]
Window inference

• Assumption 1: A significant relationship will be found to be significant in at least one window intersecting the true window

• Assumption 2: Significant relationships are a small proportion of overall set tested

• Claim: Iteratively perturbation of windows in a greedy way converges to true windows
• Key assumptions
  • Stationarity, no latent confounders
• Main advantages
  • Exact inference, time window (vs. lag), complex relationships
• Complexity: $O(N^3T)$
More on $\varepsilon_{\text{avg}}$: set $X$

- What is compared against?
  - Probability raisers at any time before $e$
  - Could be causes/effects of $c$

- Why can we do this? Two cases:
  - Independent
  - Negatively correlated
What does $\varepsilon_{\text{avg}}$ mean?

• Positive: c has positive influence on e’s probability

• Negative: probability of e greater after c’s absence (c is possible negative cause of e)

• Zero: +/- may cancel, or no influence

• Small value: maybe artifact, maybe weak real cause
Properties of $\varepsilon_{\text{avg}}$

- With large sample, no causes, values are normally distributed
Example

\[ \varepsilon_S(YF, LC) = P(LC \mid YF \land S) - P(LC \mid \neg YF \land S) \]
\[ = 0.85 - 0.75 = 0.10 \]

\[ \varepsilon_S(S, LC) = P(LC \mid S \land YF) - P(LC \mid \neg S \land YF) \]
\[ = 0.85 - 0.01 = 0.84 \]
Definition - insignificant

c is an \( \varepsilon \)-insignificant cause of \( e \) if:

c is a potential cause of \( e \) and \(|\varepsilon_{\text{avg}}| < \varepsilon\)
Definition - significant

• $c$ is a just so or $\varepsilon$-significant cause of $e$ if it is a non-$\varepsilon$-insignificant potential cause

• Note that insignificant $\neq$ spurious
  • Could be spurious or have small influence
  • Just so not necessarily genuine
  • There may be missing common causes
Testing formulas in observational data

- Specify hypotheses as temporal logical formulas
- Instead of inferring model, test whether formulas are satisfied in data
  - Find formulas satisfied at each timepoint
    - Ex: \((s \land h)Ug\)
  - Calculate conditional probability across sequences using frequencies
Probability of $c$ leads-to $e$ in 1-2 time units

- Number of times satisfying $c$: 3
- Number of times satisfying leads-to formula: 2
- Probability $= \frac{2}{3}$

$Satisfied\ by\ trace\ if\ p \leq \frac{2}{3}$
Another example

- To determine significance, need to calculate probabilities

\[
f = c \land d \overset{\geq s, \leq t}{\implies} e
\]

- Count frequency, sum over all traces

\[
\sum_{p \in P} \left( \text{# times satisfying } f \right) / \sum_{p \in P} \left( \text{# times satisfying } c \land d \right)
\]
Congestive heart failure

Temporal Causality of Social Support in an Online Community for Cancer Survivors

Ngot Bui, John Yen, and Vasant Honavar
College of Information Sciences and Technology

**ABSTRACT**

The Islamic St insurgent group prominence with paper, we press group using a datary activity sur it (including Iraq We combine idea soning to mine f evidence of caus ISIS vehicle-bour activity in Syria air strikes, and It serve as indicators of sp and arrests.

To the observed benefits have been lacking. This paper reports results of a study that seeks to address this gap by discovering temporal causality of the dynamics of sentiment change (on the part of the thread originators) in CSN. The resulting accounts offer new insights that the designers, managers and moderators of an online community such as CSN can utilize to facilitate and enhance the interactions so as to better meet the social support needs of the community participants. The proposed methodology also has broad applications in the discovery of temporal causality from big data.

**Fig. 2: Probabilistic Kripke Structure for CSN**

```
armedAtkSpike(Syria, σ)   
```

Comparing to consider in measures in [10] – age increase in considered approach, we are following:

- Indicative of operations in Iraq
- VBIED
- VBIED (p = 1.0,
Mining for Causal Relationships: A Data-Driven Study of the Islamic State

Andrew Stanton, Amanda Thart, Ashish Jain, Priyank Vyas, Arpan Chatterjee, Paulo Shakarian
Arizona State University
Tempe, AZ 85287
{dstanto2, althart, ashish.jain.1, pvyas1, achatt14, shak} @asu.edu

Table 6: Causal Rules for Spikes in VBIED Operations in Syria

<table>
<thead>
<tr>
<th>No.</th>
<th>Precondition</th>
<th>$\epsilon_{avg}$</th>
<th>$p$</th>
<th>$p^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>armedAtkSpike($Iraq, \sigma$) $\land$ indirFireSpike($Iraq, 2\sigma$)</td>
<td>0.92</td>
<td>1.00</td>
<td>0.20</td>
</tr>
<tr>
<td>6.</td>
<td>armedAtkSpike($Iraq, \sigma$) $\land$ indirFireSpike($Iraq, 2\sigma$) $\land$ VBIED($Baghdad$)</td>
<td>0.92</td>
<td>1.00</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Chronic diseases are managed primarily by individuals — adjusting insulin dosing, making and logging food choices.

Hospital data is episodic:

- Is patient on medication between visits? Are they following dietary recommendations?

Outpatient data brings more uncertainty.
Partial solution: probabilistic observations

- Issues
  - Error in measurements
  - Delay
  - Inconsistent timescales

- Strategy
  - Instead of true/false, use probability of event at each time
Discretization
Adding uncertainty to causal inference

\[ P(e|c, x) = \frac{\sum_t ecx}{\sum_t cx} \]

\[ P(e|c, x) = \frac{\sum_t P(e, c, x)}{\sum_t P(c, x)} \]
detail with time windows

$$P(e|c \land x) = \frac{\sum_{t \in T} P(c_t) P(x_{i \land \ldots \land j}) P(e_{k \land \ldots \land l})}{\sum_{t \in T} P(c_t) P(x_{i \land \ldots \land j})}$$

$$P(e|\neg c \land x) = \frac{\sum_{t \in T} P(\neg c_{g \land \ldots \land h}) P(x_t) P(e_{k \land \ldots \land l})}{\sum_{t \in T} P(\neg c_{g \land \ldots \land h}) P(x_t)}.$$
Causes of changes in glucose

Cohort: 17 subjects with T1DM

Sensor data (collected for >72 hours)

- Glucose values
- Insulin dosage
- Activity
- Sleep stage
- Heart rate
- Temperature

With N. Heintzman (Dexcom) [http://dial.ucsd.edu/what-we-do.php](http://dial.ucsd.edu/what-we-do.php)
Results

very vigorous exercise leads to hyperglycemia (fdr < .01) in 5-30 minutes
  • Found using both HR (anaerobic activity zone) and METs

  • Supported by literature (Marliss and Vranic, 2002; Riddell and Perkins, 2006)
Explanations come from:
- Need formal representation for algorithms
- Type ≠ Token
  - Cannot assume observations will exactly fit what is known
- Data
  - Frequently missing
  - Time of event ≠ time event is recorded
Automating explanation with simulation

• Why did Joe’s glucose drop in the afternoon?

• What if this instance differs in timing from what usually happens?

• What happens if our knowledge of causes is incomplete?
Example

Frank uses a CGM to help manage type 1 diabetes.

He goes for a run first thing in the morning.

He has lunch at 12pm, with a normal insulin bolus.

Several hours later, he has unexpected low blood sugar.

Did the morning run cause the low blood sugar?
• Goals for explanation

• Find causes of specific events automatically (no human in the loop)

• Find causes of when, whether and how events occur

• Approach: simulation to answer counterfactual queries

Counterfactual vs. Actual Distributions

- $P(\bullet | \neg A)$ is the **counterfactual distribution** of $A$

  $$P(B | \neg A) = \frac{P_{t_A-\tau}(B \cap \neg A)}{P_{t_A-\tau}(\neg A)}$$

- $P(\bullet | A)$ is the **actual distribution** of $A$

  $$P(B | A) = \frac{P_{t_A-\tau}(B \cap A)}{P_{t_A-\tau}(A)}$$

Three Types of Explanation

**probability:**

B **because of** A iff
\[ P(B|A) >> P(B|\neg A) \]

B **despite** A iff
\[ P(B|A) << P(B|\neg A) \]

**timing:**

B **hastened by** A iff
\[ E[t_b|A] << E[t_b|\neg A] \]

B **delayed by** A iff
\[ E[t_b|A] >> E[t_b|\neg A] \]

**intensity:**

B **intensified by** A iff
\[ E[m_b|A] >> E[m_b|\neg A] \]

B **attenuated by** A iff
\[ E[m_b|A] << E[m_b|\neg A] \]

\[ t_b = \text{time of } B \text{ occurring} \]

\[ m_b = \text{intensity (manner) of } B \text{ occurring} \]
Hastening

• 6->8, then 8->P
• but 7->8 is a more reliable backup
• probability raising finds “8->P despite 6->8”
• but by analyzing timing we find “6->8 hastened 8->P”
Diabetes Simulation

- Run or Lunch alone would not have caused Hypoglycemia (see counterfactual dists)
- Yet together they explain the Hypoglycemia (see actual distribution)
- We see beyond the most recent event (Lunch)
- We can measure quantitative strength of effect in mg/dL: $E[\text{glu} | R] - E[\text{glu} | \neg R]$
Three Types of Explanation

**probability:**
- B **because of** A iff
  \[ P(B|A) >> P(B|\neg A) \]
- B **despite** A iff
  \[ P(B|A) << P(B|\neg A) \]

**timing:**
- B **hastened by** A iff
  \[ E[t|A] << E[t|\neg A] \]
- B **delayed by** A iff
  \[ E[t|A] >> E[t|\neg A] \]

**intensity:**
- B **intensified by** A iff
  \[ E[m|A] >> E[m|\neg A] \]
- B **attenuated by** A iff
  \[ E[m|A] << E[m|\neg A] \]

\( t_B = \) time of B occurring

\( m_B = \) intensity (manner) of B occurring
For next week

- Discussion paper changed:


- Bring questions for midterm review!!