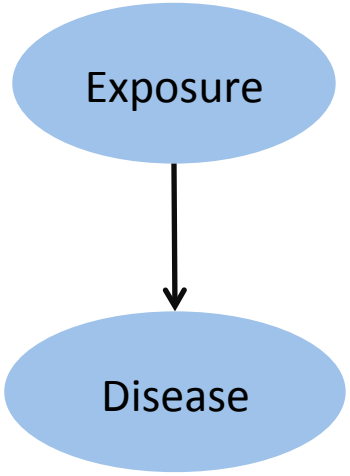
Causal Inference: prediction, explanation, and intervention

Lecture 6: Causality in time series (part I)

Samantha Kleinberg
samantha.kleinberg@stevens.edu

Recap



Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial

Leonard Leibovici

Department of Medicine, Beilinson Campus, Rabin Medical Center, Petah-Tiqva 49100, Israel

Leonard Leibovici *professor*

leibovic@post.tau.ac.il

BMJ 2001;323:1450–1

Abstract

Objective To determine whether remote, retroactive intercessory prayer, said for a group of patients with a bloodstream infection, has an effect on outcomes.
Design Double blind, parallel group, randomised controlled trial of a retroactive intervention.
Setting University hospital.
Subjects All 3393 adult patients whose bloodstream infection was detected at the hospital in 1990-6.
Intervention In July 2000 patients were randomised to a control group and an intervention group. A remote, retroactive intercessory prayer was said for the well being and full recovery of the intervention group.
Main outcome measures Mortality in hospital, length of stay in hospital, and duration of fever.
Results Mortality was 28.1% (475/1691) in the intervention group and 30.2% (514/1702) in the control group (P for difference = 0.4). Length of stay in hospital and duration of fever were significantly shorter in the intervention group than in the control group (P = 0.01 and P = 0.04, respectively).
Conclusions Remote, retroactive intercessory prayer said for a group is associated with a shorter stay in hospital and shorter duration of fever in patients with a bloodstream infection and should be considered for use in clinical practice.

were included in the study. Bloodstream infection was defined as a positive blood culture (not resulting from contamination) in the presence of sepsis.

In July 2000 a random number generator (Proc Uniform, SAS, Cary, NC, USA) was used to randomise the patients into two groups. A coin was tossed to designate the intervention group. A list of the first names of the patients in the intervention group was given to a person who said a short prayer for the well being and full recovery of the group as a whole. There was no sham intervention.

Three primary outcomes were compared: the number of deaths in hospital, length of stay in hospital from the day of the first positive blood culture to discharge or death, and duration of fever. Patients were defined as having fever on a specific day if one of three temperature measurements taken on that day showed a temperature of >37.5°C.

The χ^2 test was used to test for the significance of the results shown in the tables. As most of the continuous variables did not have a normal distribution, the Wilcoxon rank sum test was used for comparisons.

Results

Of 3393 patients with a bloodstream infection, 1691 patients were randomised to the intervention group and 1702 to the control group. No patients were lost to

Leibovici, L. (2001). Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: Randomised controlled trial. *Bmj*, *323*(7327), 1450-1451.

Paper proves power of statistics, not prayer

Shehan Hettiaratchy, fellow, transplantation biology

Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA ; Email: shehan.hettiaratchy@tbrc.mgh.harvard.edu

Carolyn Hemsley, Wellcome Trust fellow in microbiology

Author information ► Copyright and License information ►

EDITOR—It was very brave of both Leibovici and the *BMJ* to publish this paper and be prepared for the criticism from the outraged masses.[1-1] The idea that retroactive intercessory prayers could have an influence on the outcome of septicaemia is intriguing and challenges our notions of cause and effect. If it is true, however, this is not the paper to prove it.

The data on the most significant finding, length of stay, seem to be skewed by a few abnormally high results in the control group. This is shown by the fact that the median length of stay is the same in both groups but the maximum length of stay in the control group is twice that in the intervention group. This may represent a type I statistical error, despite the large sample size. From a cynical standpoint, it is a shame that God cannot save your life but might get you out of hospital a few days earlier. Either way, it was a thought provoking paper but may just prove the power of statistics, not of prayer.

References

Go to:

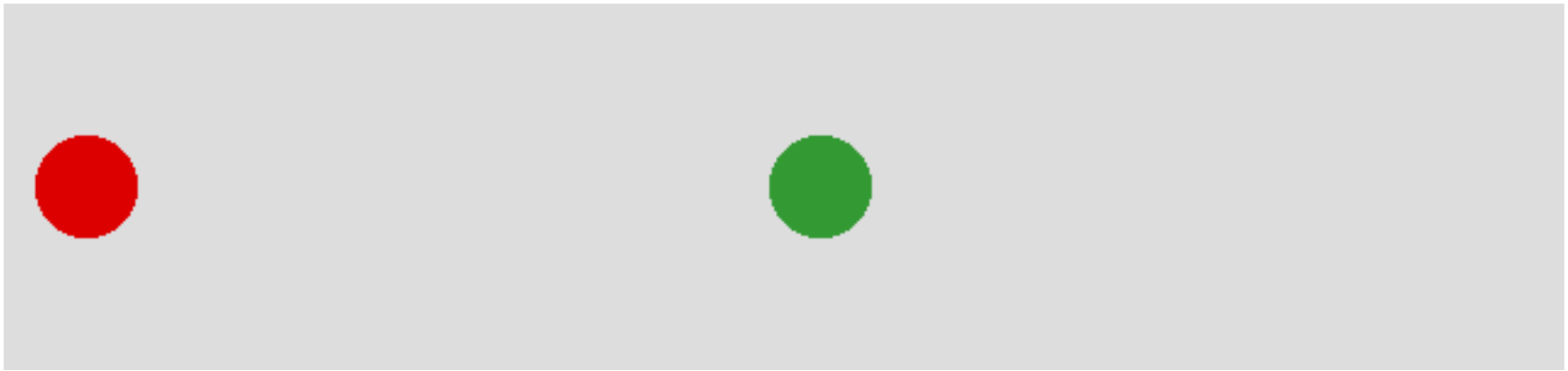
1-1. Leibovici L. Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial. BMJ. 2001;323:1450–1451. . (22 December.) [PMC free article] [PubMed]

Contrast

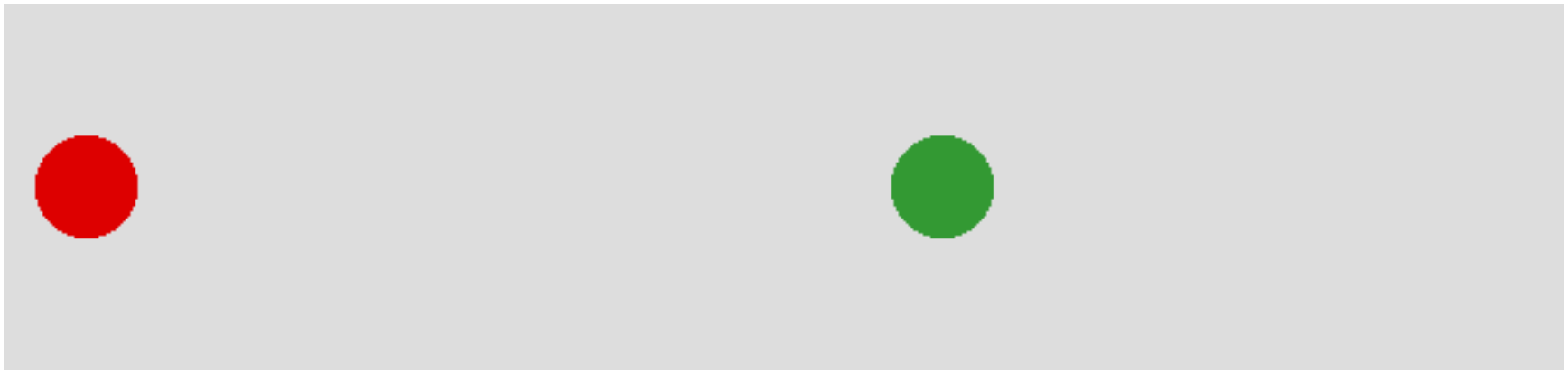
- Exposure to someone with chickenpox Tuesday, illness Friday
- Exposure to someone with chickenpox in September, illness in November

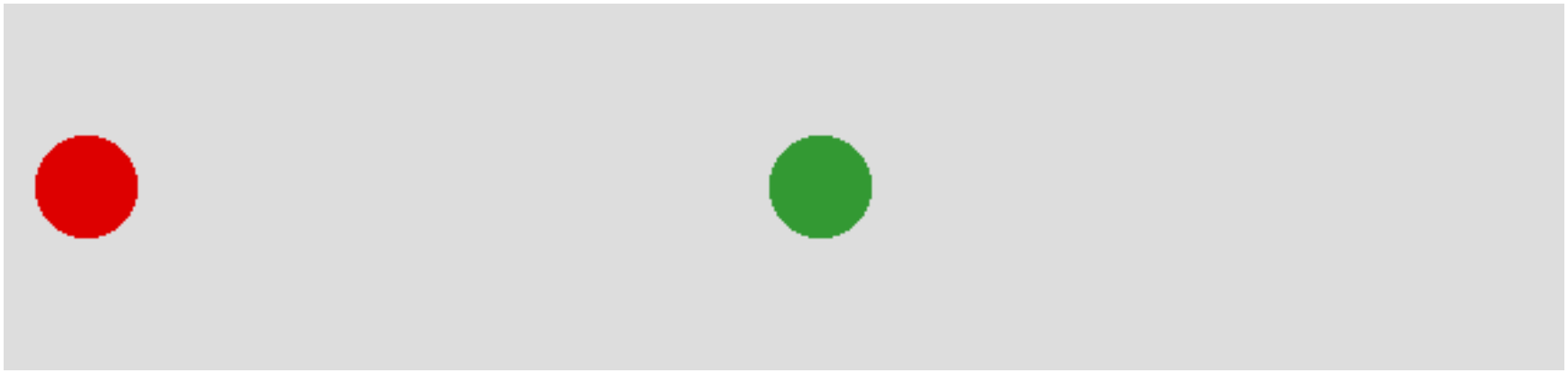


(a) Simultaneous motion



Following videos from: Scholl, B. J., & Tremoulet, P. (2000). Perceptual causality and animacy. ***Trends in Cognitive Sciences****, 4(8)*, 299-309.







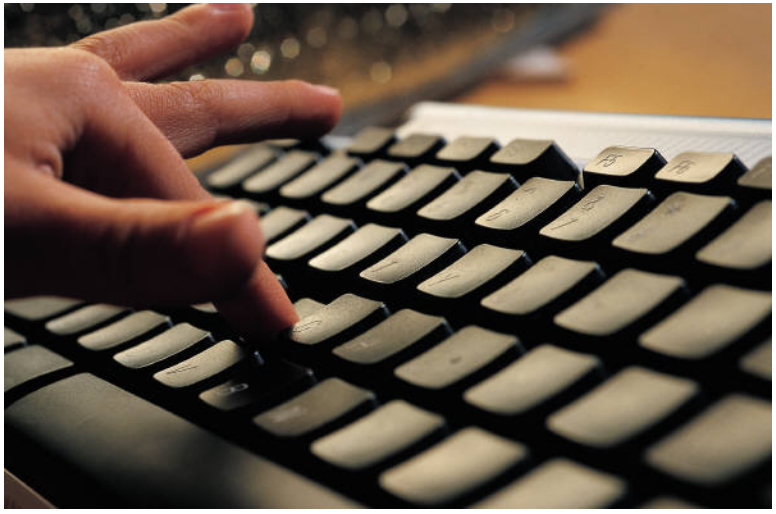
(b) Launching



(c) Delayed launching

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)

Buehner MJ, McGregor S (2006) Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. Thinking & reasoning 12(4):353-378.

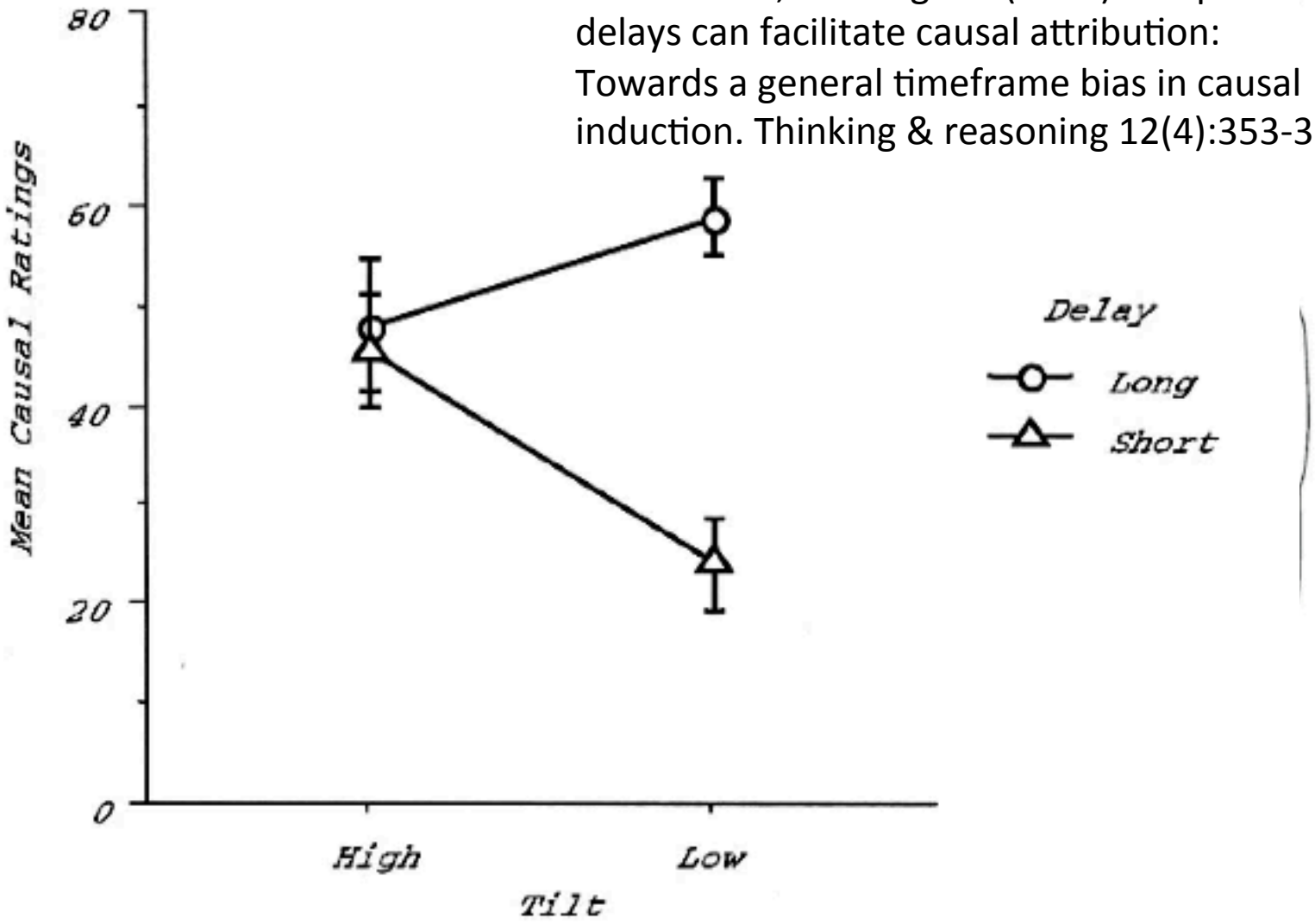


Figure 3. Mean causal ratings from Experiment 1.

Figur ...r on the right.

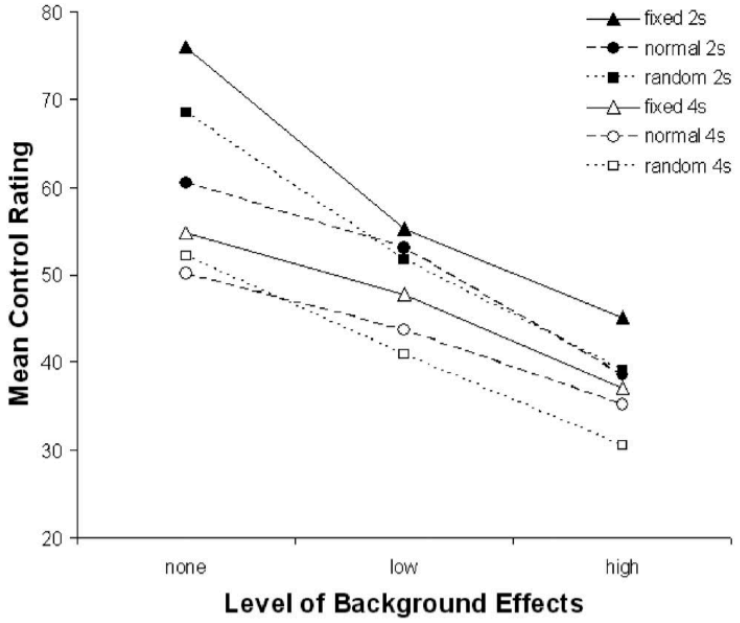
- Time used to learn causality (but can incorrectly override covariation)
- More consistent vs variable timing

Temporal Predictability Facilitates Causal Learning

W. James Greville and Marc J. Buehner
Cardiff University

Temporal predictability refers to the regularity or c
When encountering repeated instances of causes and
temporal intervals. Where this interval is constant it
follow from the cause. In contrast, interval variab
investigated the extent to which temporal predictabili
causal learning. The authors demonstrated that (a)
consistently judged as stronger than those with vari
decline as a function of temporal uncertainty, and (c) t
learning time. The results therefore clearly indicate th
ery. The authors considered the implications of th
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

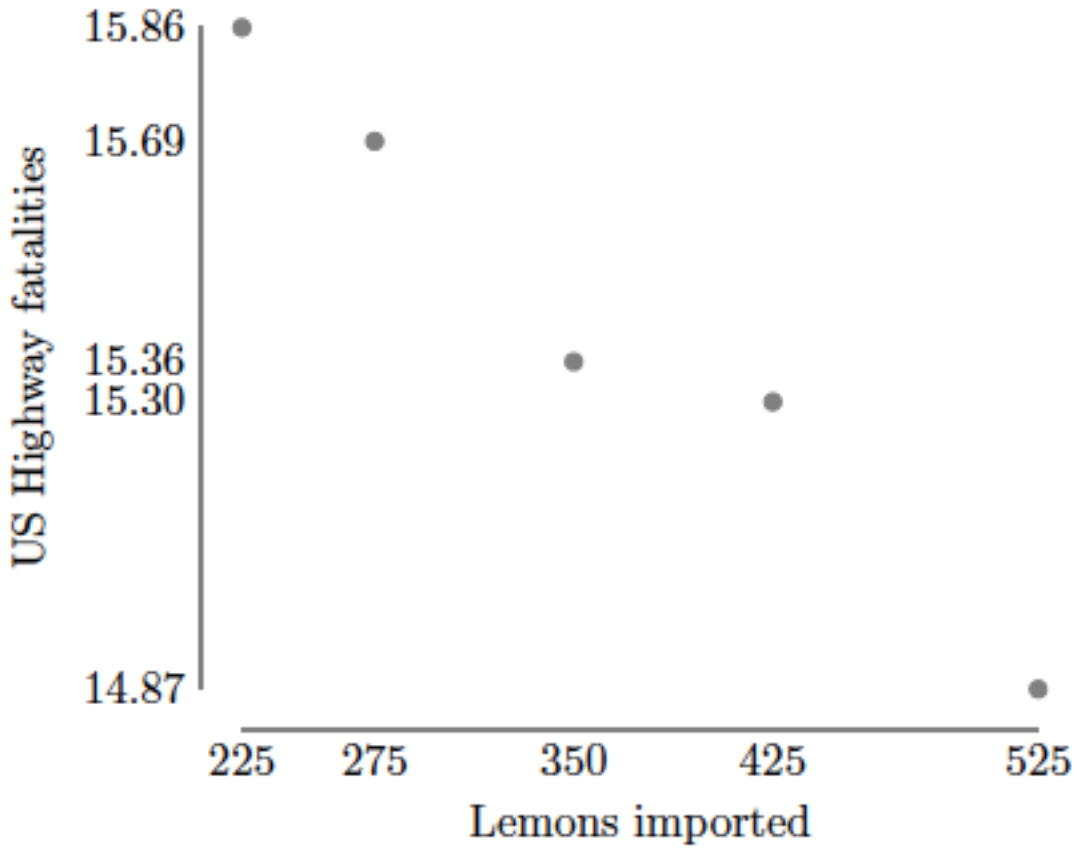
Abstract: The close relation between time and causality is undisputed, but there is a paucity of research on how people use temporal information to inform their causal judgments. Experiment 1 examined the effect of delay variability on causal judgments, and whether participants were sensitive to the 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.

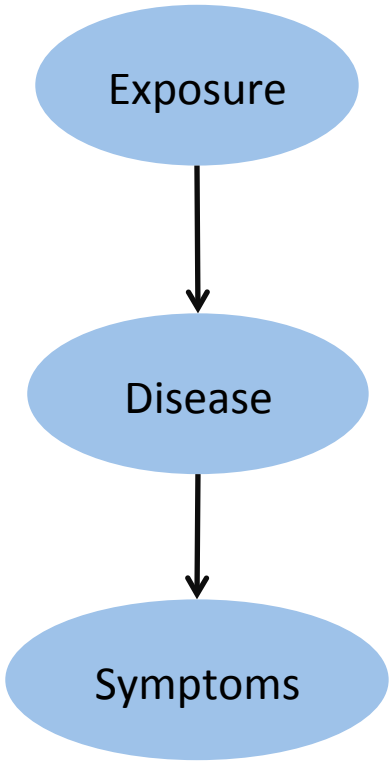
But...

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

Nonstationarity



Time in graphical models



First-order Markov process

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \ldots \rightarrow X_t$$

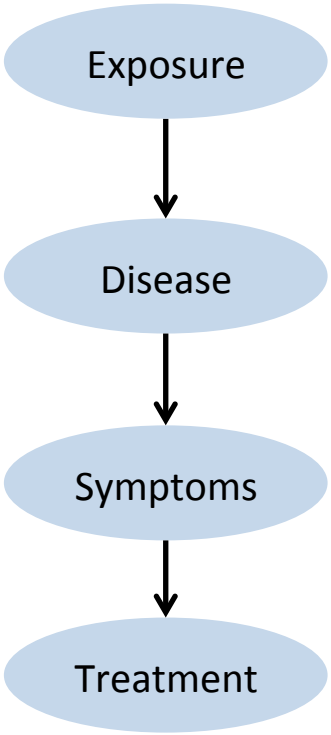
- Edges only between i and i+1

$$P(X_t \mid X_0 \ldots X_{t-1}) = P(X_t \mid X_{t-1})$$

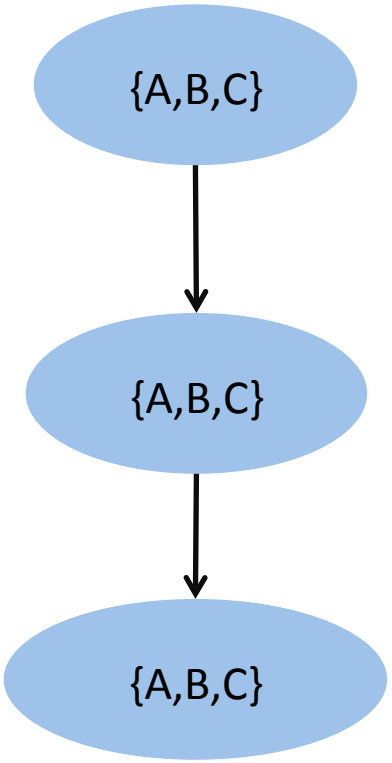
- Future and past are independent conditioned on present

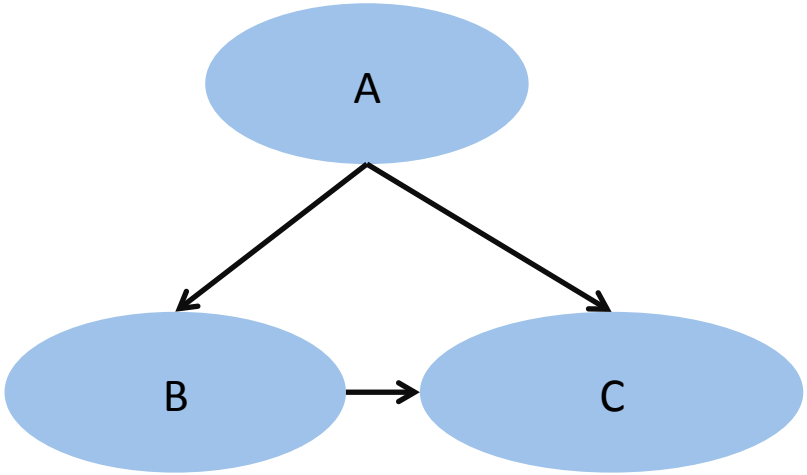
Markov processes

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

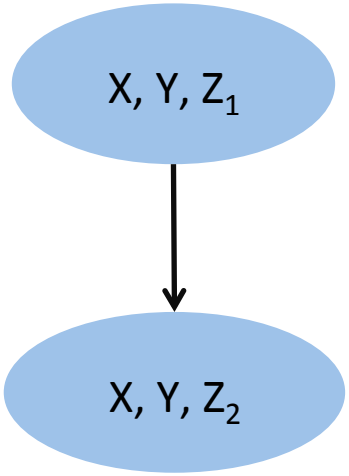


States vs. variables





Example

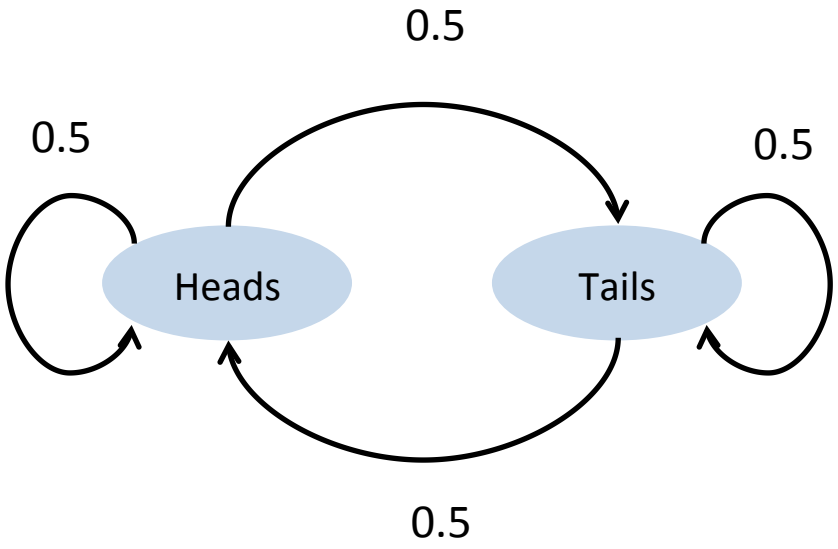


XYZ_1	XYZ_2	P
111	111	0.01
111	110	0.1
111	101	0.03
	…	
000	001	0.04
000	000	0.5

Terminology

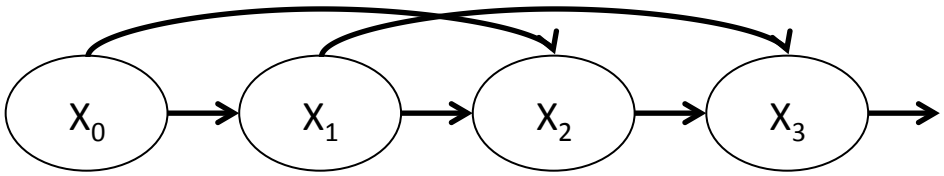
Markov process: stochastic process where transition probabilities dependent only on current state

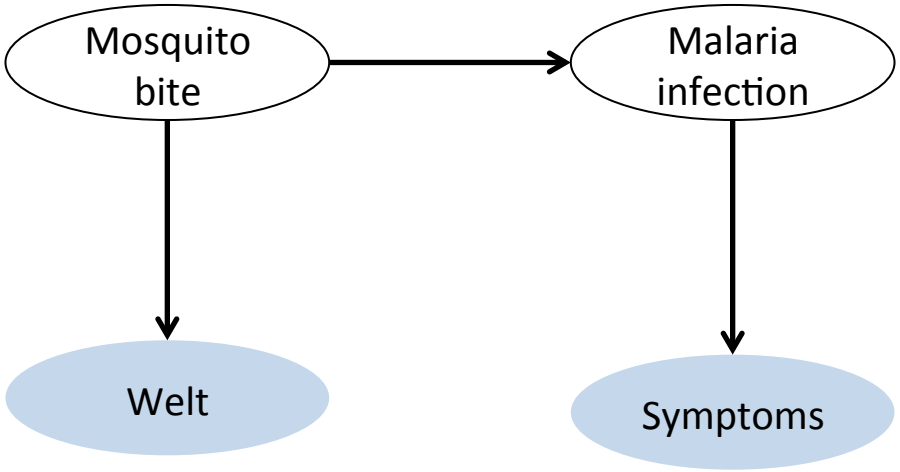
Each node is a state (collection of variable assignments)

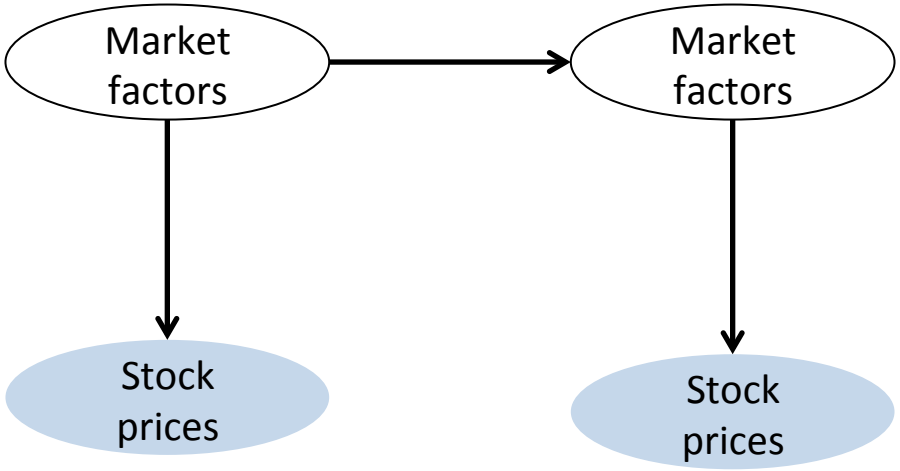


General

- First order: t depends on t-1
- Second order: t depends on t-1 and t-2

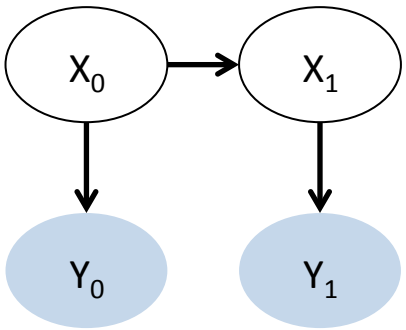


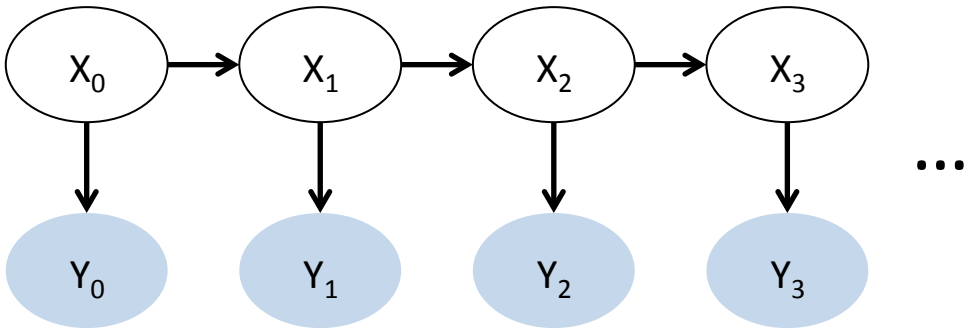

Mosquito bite
Malaria infection
Welt
Symptoms


Market factors
Market factors
Stock prices
Stock prices

Hidden Markov Model

$$P(Y_t \mid X_{0:t-1}) = P(Y_t \mid X_{t-1})$$



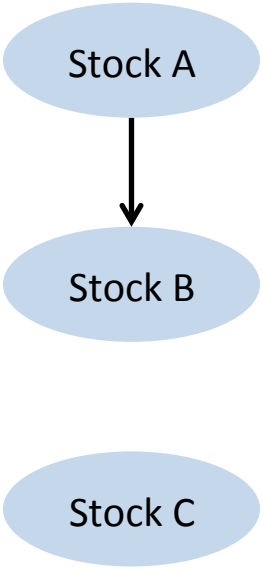


Dynamic Bayesian networks

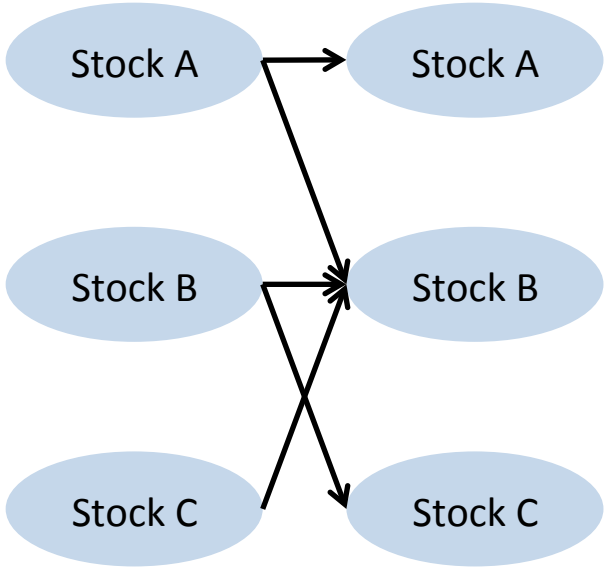
- "Dynamic" because they include time – not because they necessarily change over time!
- Generalization of HMM (HMMs are a subset of DBNs)

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

Example

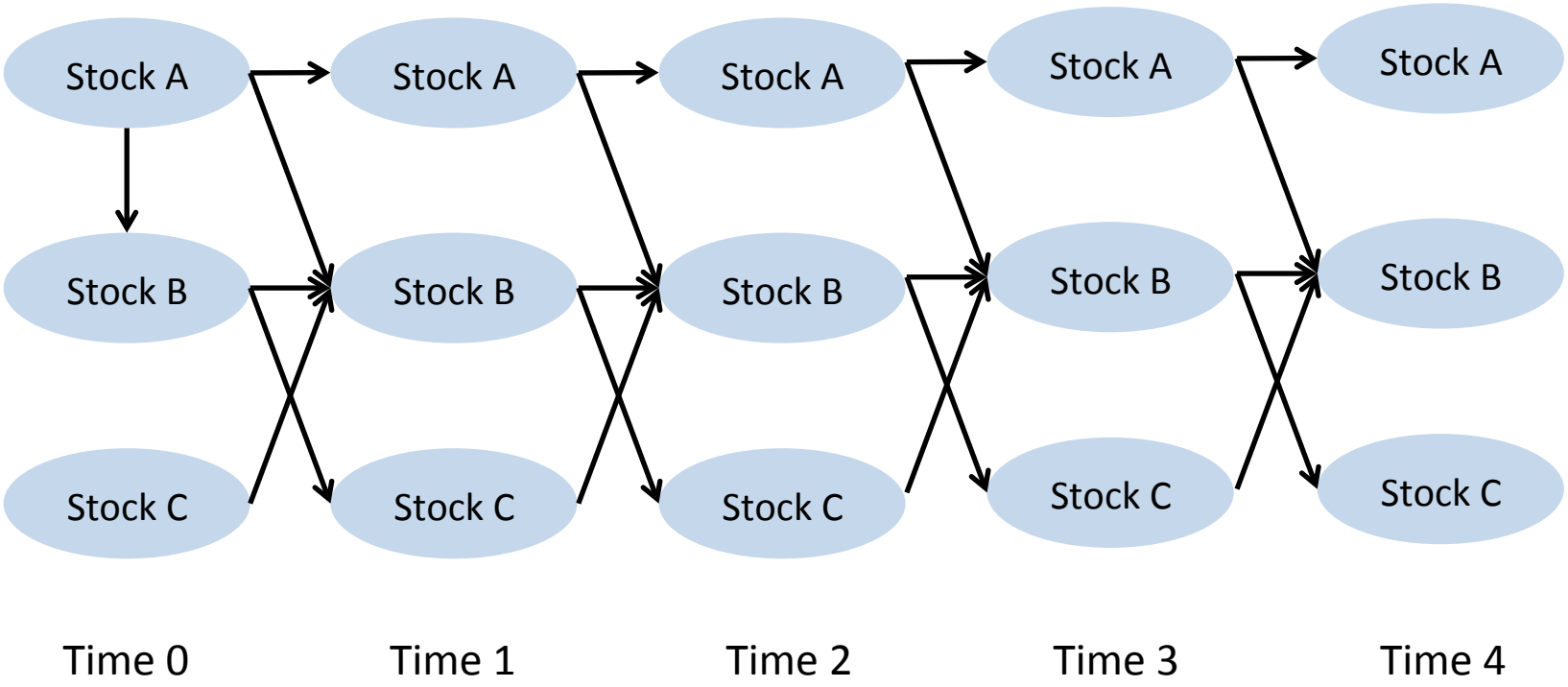


Initial



Transition from t to t+1

Unrolled


Stock A
Stock B
Stock C
Stock A
Stock B
Stock C
Stock A
Stock B
Stock C
Stock A
Stock B
Stock C
Stock A
Stock B
Stock C
Time 0
Time 1
Time 2
Time 3
Time 4

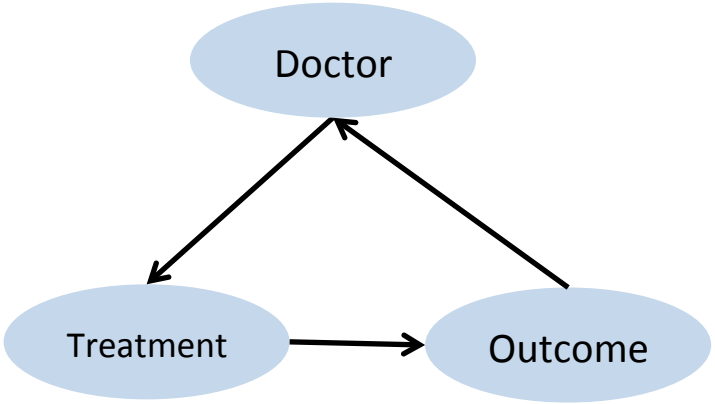
DBNs

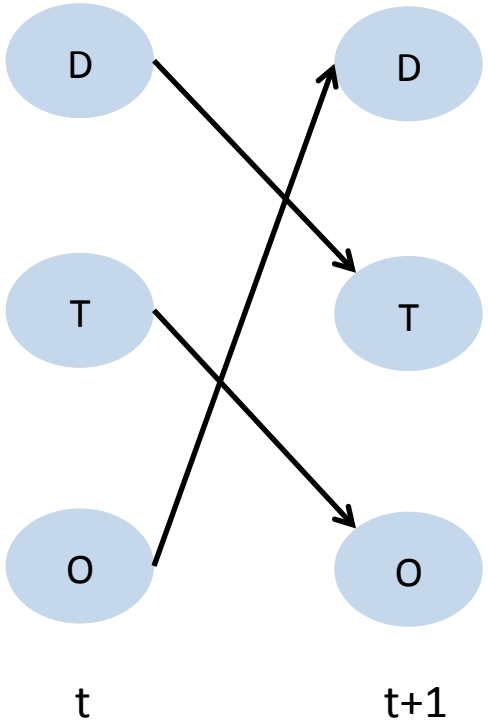
A DBN is defined by:

-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





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

	Proposed here	Robinson & Hartemink (2009)	Lèbre (2008)	Grzegorcyk et al. (2008)	Ko et al. (2007)
Score	Marginal Likelihood	Marginal Likelihood	Marginal Likelihood	Marginal Likelihood	BIC
Change-points	node specific	whole network	node specific	whole network	node specific
Structure constant	Yes	No	No	Yes	Yes
Data format	Continuous	Discrete	Continuous	Continuous	Continuous
Latent variables	Change-point process	Change-point process	Change-point process	Free allocation	Free allocation

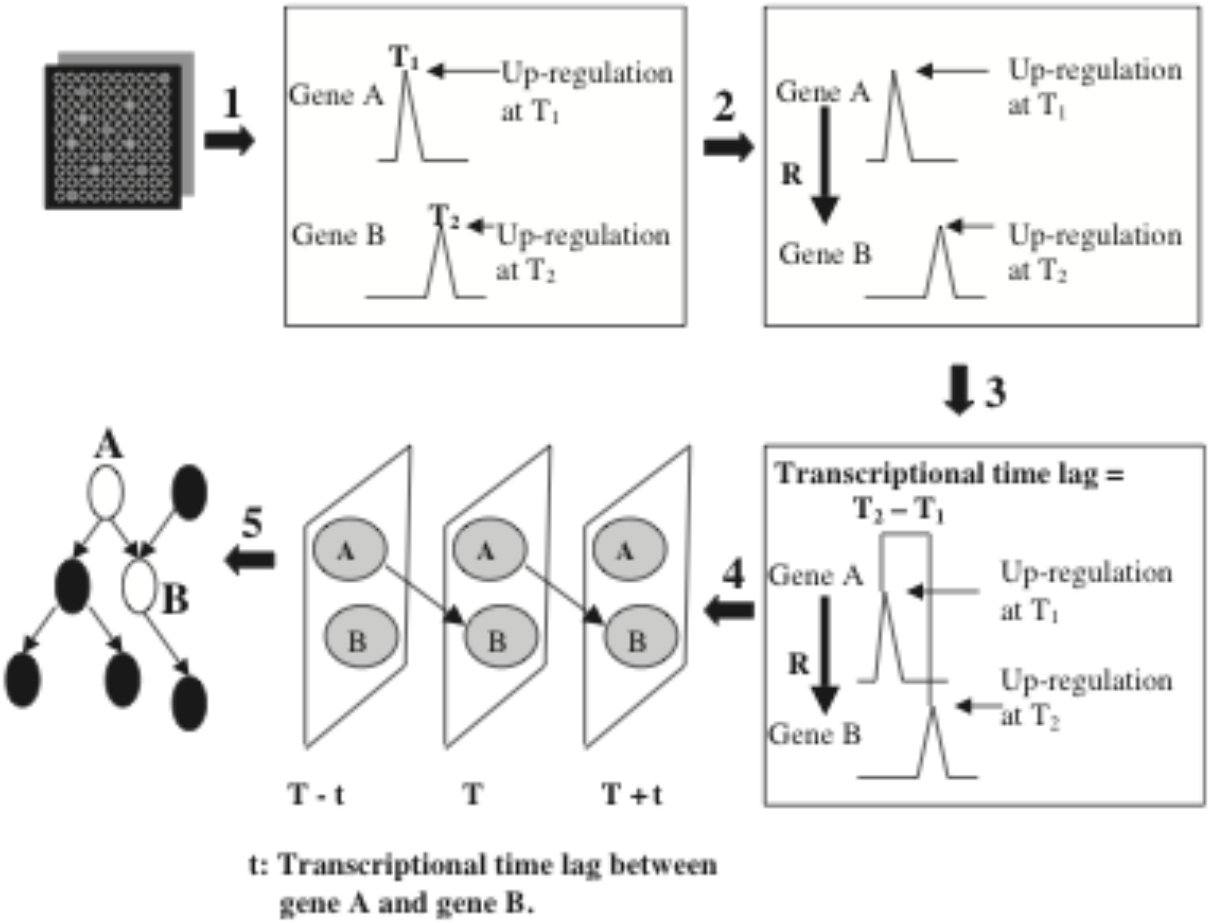
Grzegorczyk, M., & Husmeier, D. (2009). Non-stationary continuous dynamic bayesian networks. *Advances in Neural Information Processing Systems (NIPS)*, *22*, 682-690.

Some software

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

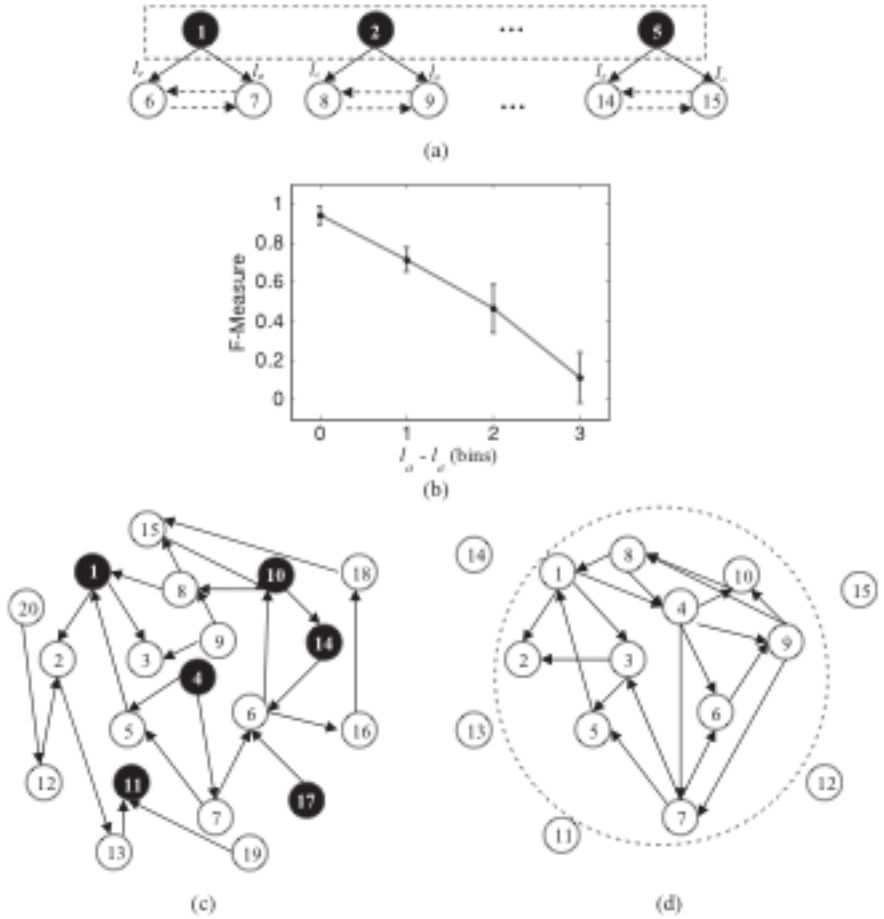
More: http://www.cs.ubc.ca/~murphyk/Software/bnsoft.html

Application: gene regulatory networks



Zou, M., & Conzen, S. D. (2005). A new dynamic bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, *21*(1), 71.

Application: neuronal connectivity



Eldawlatly, S., Zhou, Y., Jin, R., & Oweiss, K. G. (2010). On the use of dynamic bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural Computation*, *22*(1), 158-189

References

- Ghahramani Z (1998) Learning dynamic Bayesian networks. Adaptive Processing of Sequences and Data Structures:168-197.
- Murphy K (2002) Dynamic Bayesian networks: representation, inference and learning. PhD Thesis, University of California, Berkley.

Logic-based

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

The following slides are from:
Kleinberg, S (2013). *Causality, Probability and Time*, Cambridge University Press.

Figure 8: Spikes in ISIS Suicide Operations in Iraq per Week

4.5

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

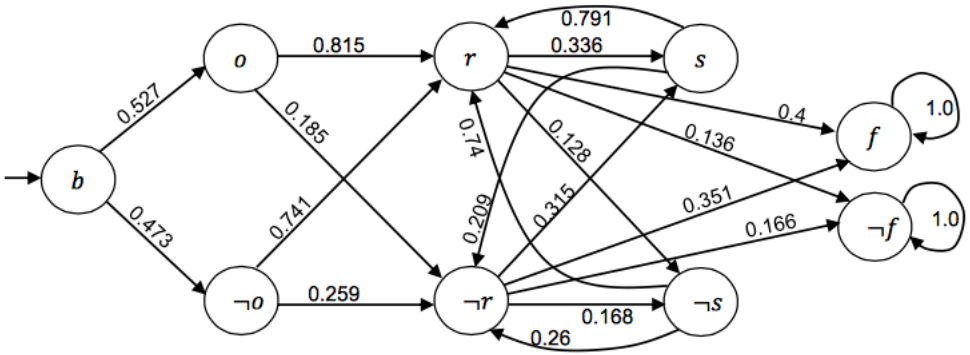


Fig. 2: Probabilistic Kripke Structure for CSN

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.

armedAtkSpike($Syria, \sigma$)

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

No.	Precondition	ϵ_{avg}	p	p^*
5.	armedAtkSpike($Iraq, \sigma$)∧ indirFireSpike($Iraq, 2\sigma$)	0.92	1.00	0.20
6.	armedAtkSpike($Iraq, \sigma$)∧ indirFireSpike($Iraq, 2\sigma$)∧ VBIED($Baghdad$)	0.92	1.00	0.20

What is a causal relationship?

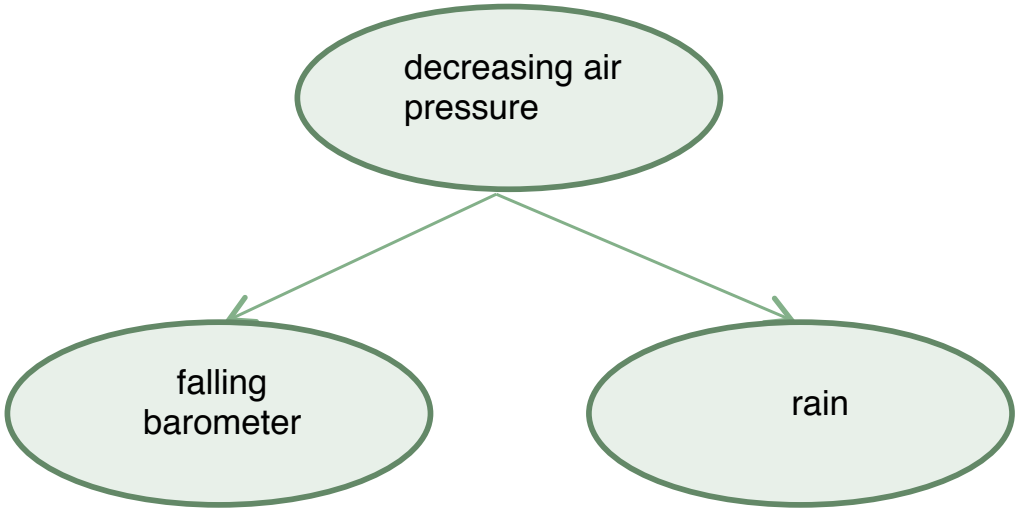
smoking and hypertension until a particular genetic mutation occurs causes CHF in 1-2 years with probability p

$$(s \wedge h) U g \overset{\geq 1, \leq 2}{\underset{\geq p}{\sim\!\!\rightarrow}} CHF$$

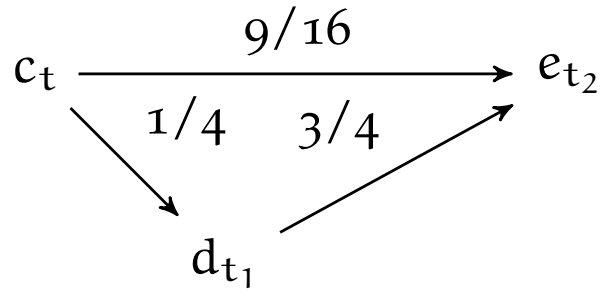
*made up relationship + timing

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

Which causes are significant?

- Main idea: looking for better explanations for the effect
- Calculate impact of cause on effect's probability - after accounting for other factors

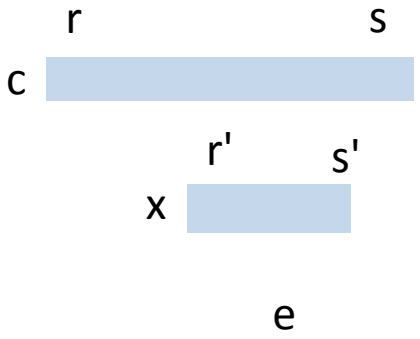
$$\varepsilon_{avg}(c,e) = \frac{\sum_{x \in X \setminus c} P(e \mid c \wedge x) - P(e \mid \neg c \wedge x)}{|X \setminus c|}$$

ε_{avg} and time

P(e|c^x) means c,x both occur and e occurs in overlapping time window

$$c \overset{\geq s, \leq t}{\underset{\geq p}{\rightsquigarrow}} e$$

$$x \overset{\geq s', \leq t'}{\underset{\geq p}{\rightsquigarrow}} e$$



More on ε_{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

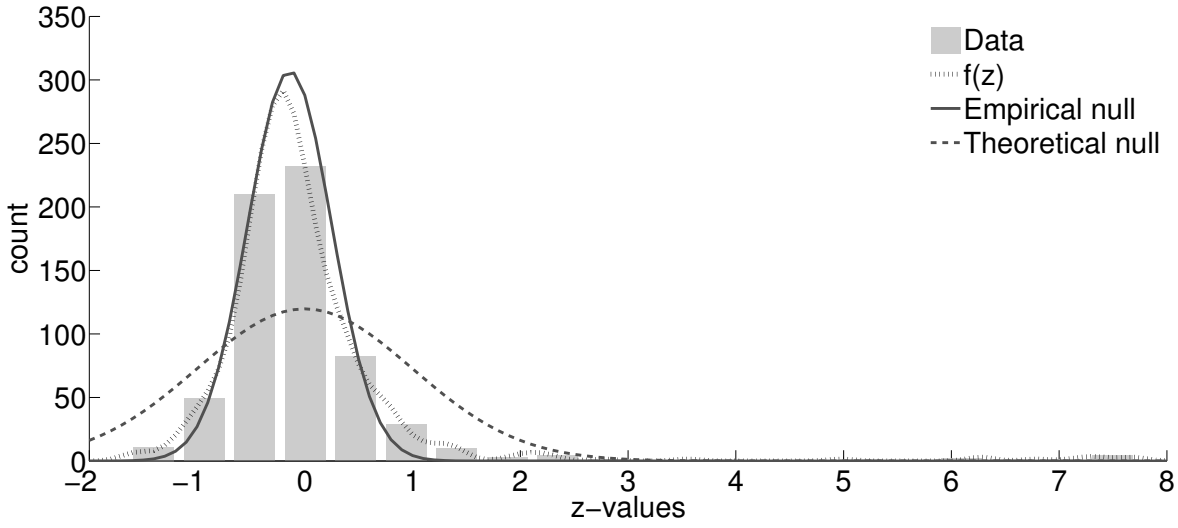
What does ε_{avg} mean?

- 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

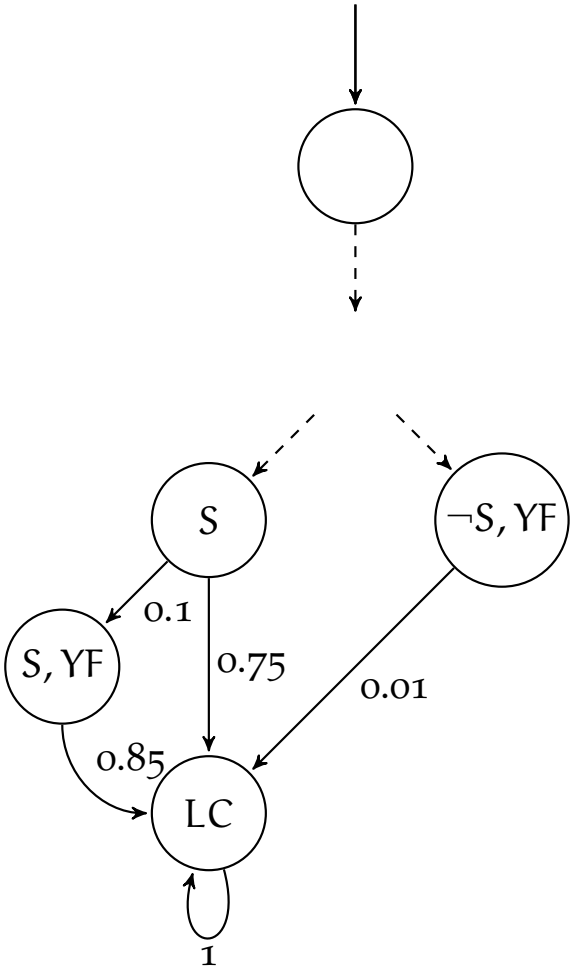
- Small value: maybe artifact, maybe weak real cause

Properties of ε_{avg}

- With large sample, no causes, values are normally distributed
- Note: not equal to zero for non-causes. Why?



Example



$$\varepsilon_S(YF, LC) = P(LC \mid YF \wedge S) - P(LC \mid \neg YF \wedge S)$$
$$= 0.85 - 0.75 = 0.10$$

$$\varepsilon_S(S, LC) = P(LC \mid S \wedge YF) - P(LC \mid \neg S \wedge YF)$$
$$= 0.85 - 0.01 = 0.84$$

Definition - insignificant

c is an **ε-insignificant** cause of e if:

c is a potential cause of e and $|\varepsilon_{avg}| < \varepsilon$

Definition - significant

- c is a **just so** or **ε-significant** cause of e if it is a non-ε-insignificant potential cause

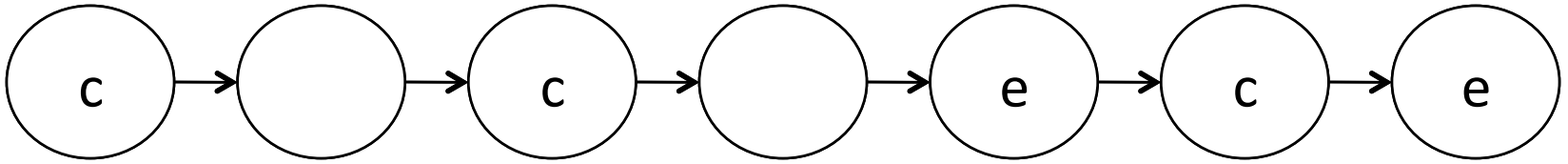
- Note insignificant ≠ spurious
  - Could be spurious or have small influence
- Just so not necessarily genuine
  - There may be missing common causes

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: $(s \wedge h)Ug$
  - 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 = 2/3



False True True

$$c \overset{\geq 1, \leq 2}{\underset{\geq p}{\rightsquigarrow}} e$$

Satisfied by trace if p≤ 2/3

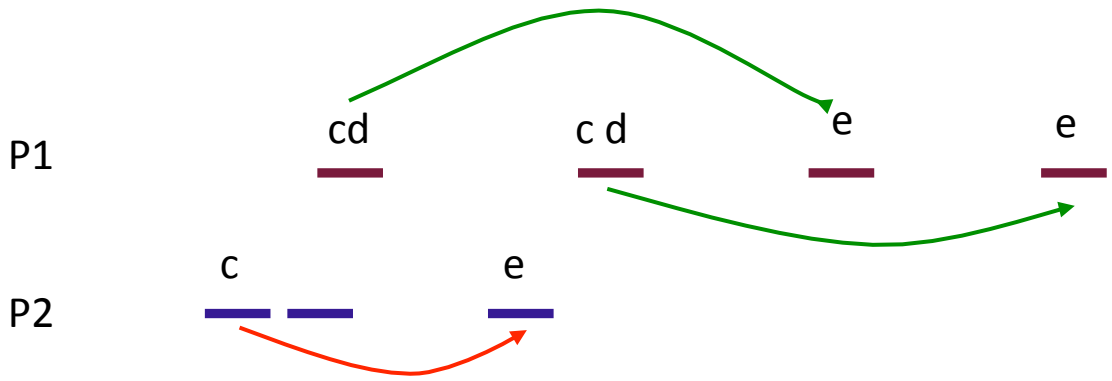
Another example

- To determine significance, need to calculate probabilities

$$f = c \wedge d \overset{\geq s, \leq t}{\rightsquigarrow} e$$

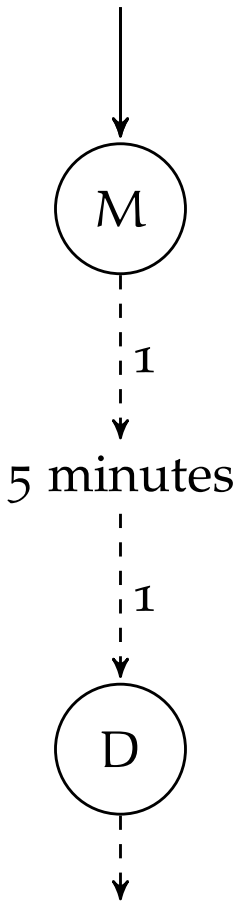
- Count frequency, sum over all traces

$$\frac{\sum_{p \in P} (\#\text{ times satisfying } f)}{\sum_{p \in P} (\#\text{ times satisfying } c \wedge d)}$$

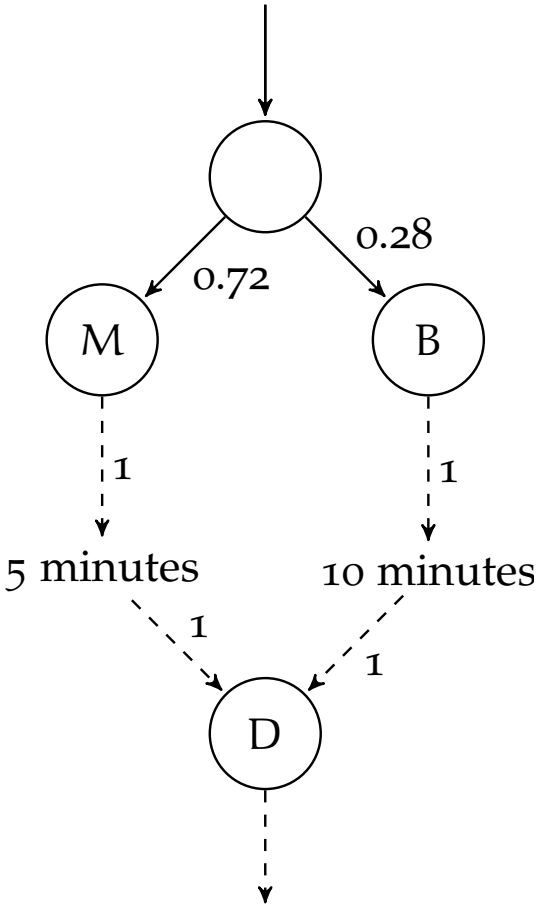


Determinism

- Events that are inevitable
- Cause that always produces effect



M necessary/sufficient



M, B sufficient, not necc

Redundant causation

- Overdetermination
  - Multiple causes that always co-occur (e.g. firing squad)
- Early preemption
- Late preemption
  - Causes both occur, one routinely preempts other

Chains

- Always finds most direct cause. Why?

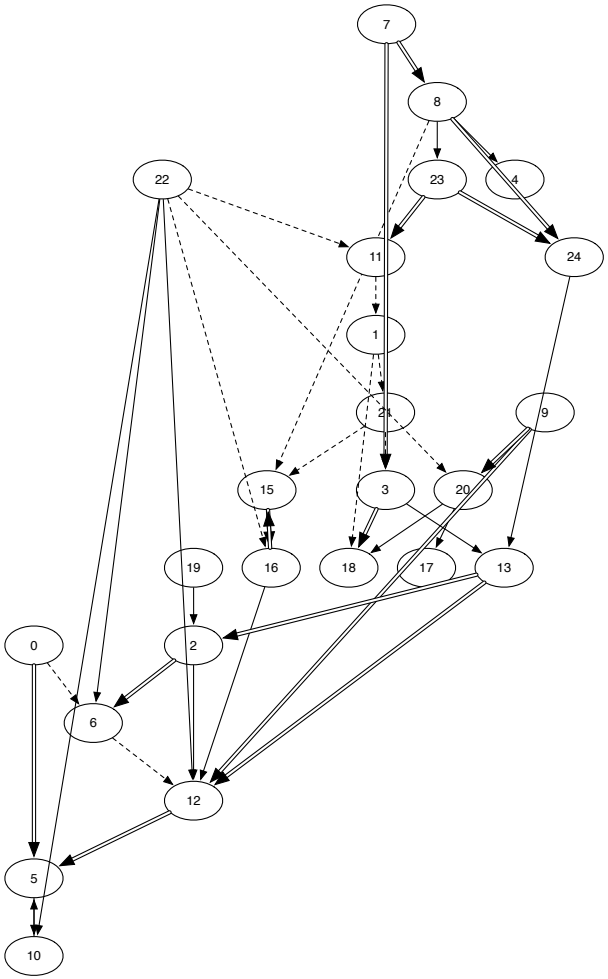
Application: simulated financial time series

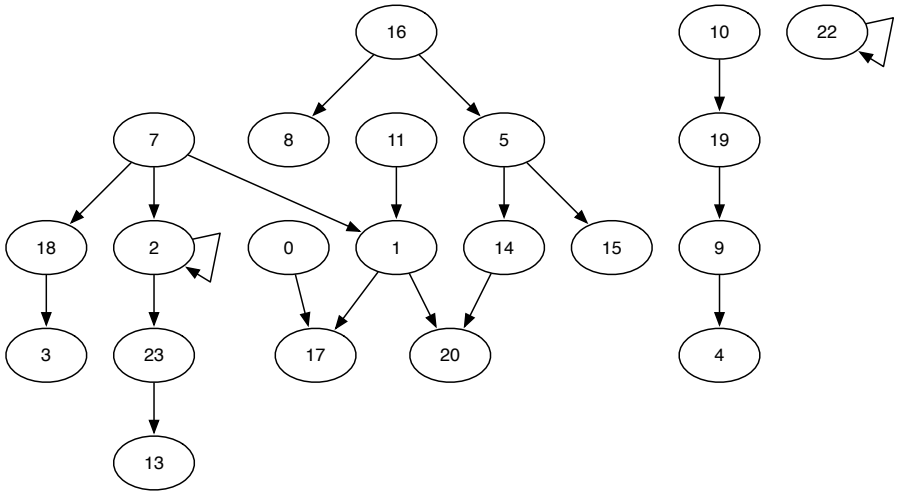
- Created synthetic data based on Fama-French factor model [Fama & French 1992]

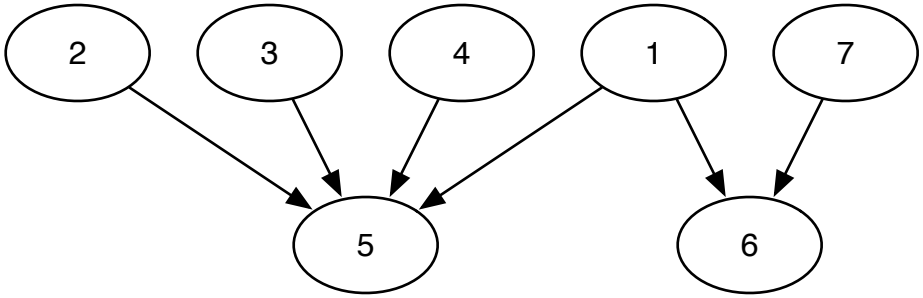
$$r_{i,t} = \sum_j \beta_{ij} f_{j,t} + \varepsilon_{i,t}$$

- 11 datasets, two time periods
  - [1] no causality
  - [2-8] randomly generated relationships, one lag
  - [9-10] randomly generated relationships, random lags
  - [11] many-to-one relationships, one lag

Patterns


7
8
22
23
4
11
24
1
21
9
15
3
20
19
16
18
17
13
0
2
6
12
5
10


16
10
22
7
8
11
5
19
18
2
0
1
14
15
9
3
23
17
20
4
13


2
3
4
1
7
5
6

Results

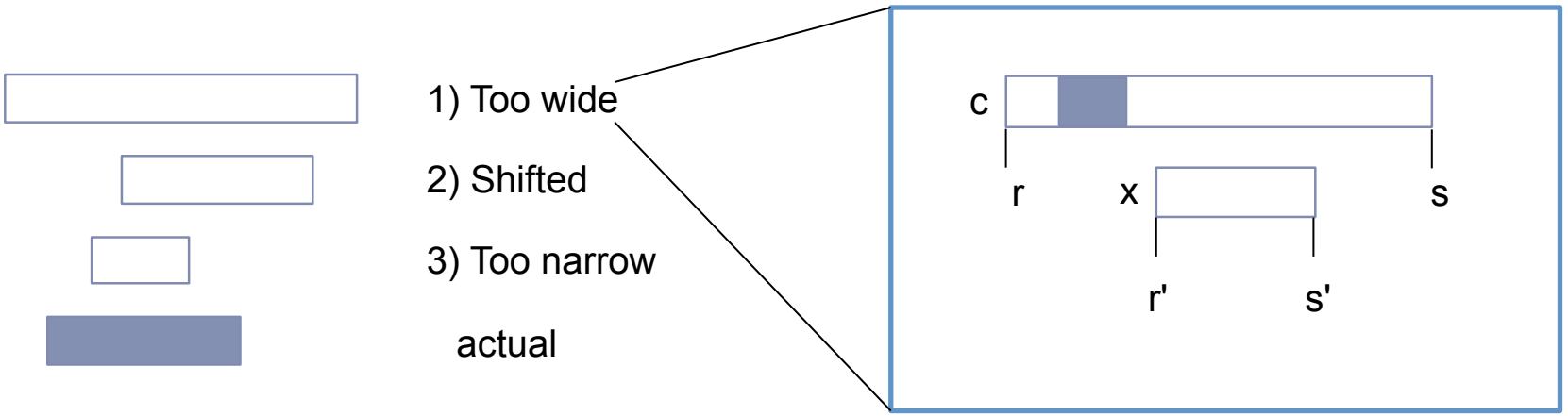
- Results for return data below
  - 22 datasets (11 types, 2 runs each)
    - Two 4000 day time periods
  - 25 stocks, 3 factors
  - Daily returns
- True positives
  - Dependency between stocks

FDR	FNR	Intersection
0.0241	0.0082	0.4714

Inferring timing

- From testing relationships between all variables and CHF in 1-2 weeks, want to ultimately infer "high AST leads to CHF in 4-10 days"
- Instead of only accepting/rejecting hypotheses, refine hypotheses

Intuition behind procedure



$$\varepsilon_{avg}(c,e) = \frac{\sum_{x \in X \setminus c} P(e \mid c \wedge x) - P(e \mid \neg c \wedge x)}{|X \setminus c|}$$

$$P(e \mid c \wedge x) = \frac{\#(c \wedge x \wedge e)}{\#(c \wedge x)}$$

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

Recap

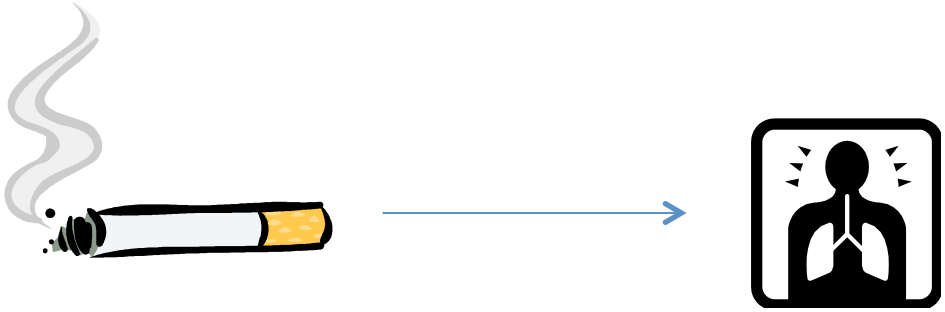
- Find significant relationships
  - Generate initial hypotheses with candidate time windows, find possible causes
  - Calculate ε_{avg} for each
  - Label possible causes with fdr > threshold as ε-insignificant, rest ε-significant
- For statistically significant relationships, iteratively perturb time window (expand, contract, shift) trying to maximize ε_{avg}
- Re-calculate ε_{avg} with newly inferred relationships

Explanation

- Where do 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

Timing

- LC= S
- LC=A





Intuition

- Type level relationships are candidates
- Strictly applying type-level relationships is problematic
- Without background knowledge that timing is a strong constraint, do not want this to strictly limit explanations

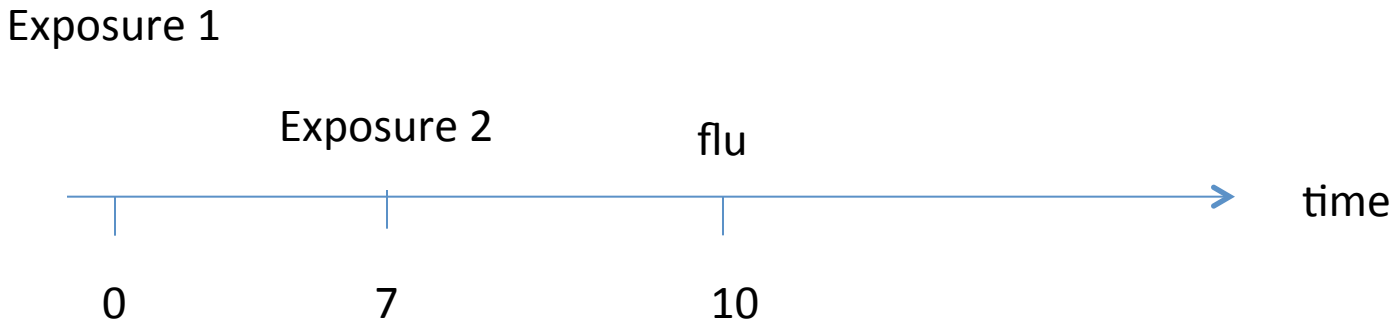
Goals

- Quantify significance of an explanation
  - Allows for incomplete knowledge
  - Incorporates temporal uncertainty
- Algorithm for finding the significance of explanations

General approach

- Infer causal relationships from data
- Observe token sequence
  - E.g. Series of medical tests or continuous recordings
- Iterate over relationships to calculate significance for each explanation
- Result is ranking of possible explanations

Significance of hypotheses



- When cause and effect occur consistent with type-level information
  - Significance = type-level significance
- When timing differs
  - Weight by difference
- When unknown if c occurs
  - Weight by probability of occurrence

Recall connecting principle!

Assessing significance

With $c \overset{\geq r, \leq s}{\underset{\geq p}{\rightsquigarrow}} e$

$$S(c_{t'}, e_t) = \varepsilon_{avg}(c, e) \times P(c_{t'} \mid V) \times f(t', t, r, s)$$

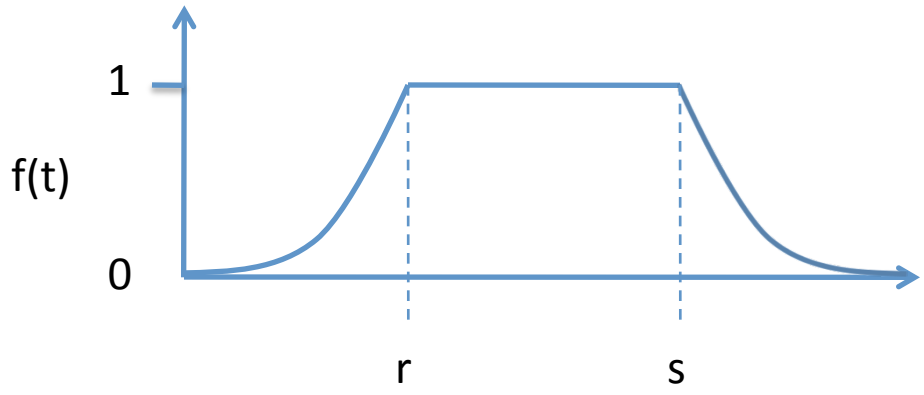
=1 if occurs at time t’

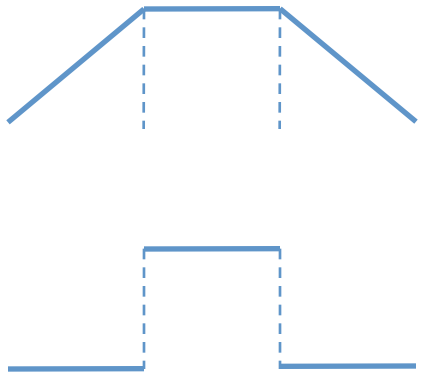
=1 if c in window [r,s]

V is observation sequence

Accounting for uncertainty

$$S(c_{t'}, e_t) = \varepsilon_{avg}(c, e) \times P(c_{t'} \mid V) \times f(t', t, r, s)$$





- Slope may be dependent on underlying mechanisms
  - i.e. when is cause still capable of producing effect
- Can use other function with same properties

Calculating P

$$S(c_{t'}, e_t) = \varepsilon_{avg}(c, e) \times P(c_{t'} \mid V) \times f(t', t, r, s)$$

- Where do probabilities come from?
  - Initial dataset
  - Model
- Conditional on token-level observations

Algorithm

1. $X \leftarrow$ type-level genuine and just so causes of e

   $\mathcal{V} \leftarrow [\mathcal{V}_0, \mathcal{V}_1 \ldots \mathcal{V}_t]$

   $EX \leftarrow \emptyset$

2. For each $x \in X$, where $x = y \rightsquigarrow^{\geqslant r, \leqslant s} e$, for $i \in [t-s, t-r]$ if $\mathcal{V}_i \models y$ the significance of y_i for e_t is $\varepsilon_{avg}(y_{r-s}, e)$ and $EX = EX \bigcup \{x\}$

3. For each $x \in X \setminus EX$, where $x = y \rightsquigarrow^{\geqslant r, \leqslant s} e$, and

$$i = \underset{j \in [0,t)}{\operatorname{argmax}}(P(y_j|\mathcal{V}) \times f(j, t, r, s)),$$

   the significance of y_i for e_t is

$$S(y_i, e_t) = \varepsilon_{avg}(y_{r-s}, e) \times P(y_i|\mathcal{V}) \times f(i, t, r, s).$$

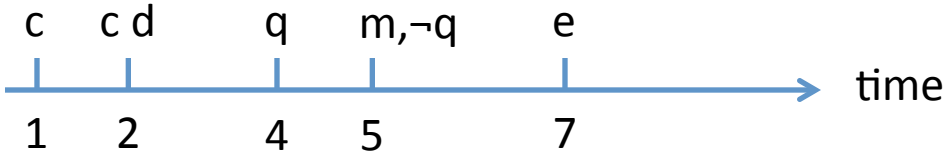
Example of procedure

- Type-level relationships (X)

$$c \wedge d \overset{\geq 3, \leq 5}{\rightsquigarrow} e \qquad \varepsilon_{avg}(c \wedge d, e)$$

$$m \overset{\geq 1, \leq 1}{\rightsquigarrow} e \qquad \varepsilon_{avg}(m, e)$$

- Token observation (V)



Example of procedure

- Type-level relationships (X)

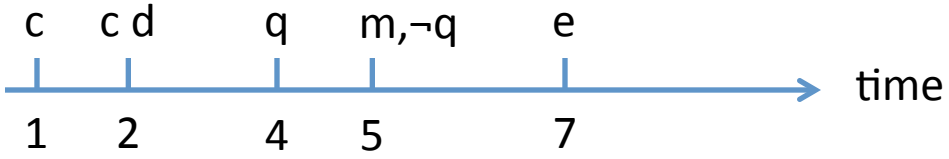
$$c \wedge d \overset{\geq 3, \leq 5}{\rightsquigarrow} e \qquad \varepsilon_{avg}(c \wedge d, e)$$

$$m \overset{\geq 1, \leq 1}{\rightsquigarrow} e \qquad \varepsilon_{avg}(m, e)$$

$$\varepsilon_{avg}(m, e) \times f(5,7,1,1)$$

$$\varepsilon_{avg}(m, e) \times P(m_6 \mid V)$$

- Token observation (V)



For next week

- Read: C. W. J. Granger (1980). Testing for Causality: A Personal Viewpoint.

- Bring questions for midterm review!!