Causal Inference:
prediction, explanation, and intervention

Lecture 6: Causality in time series (part I)

Samantha Kleinberg
samantha.kleinberg@stevens.edu

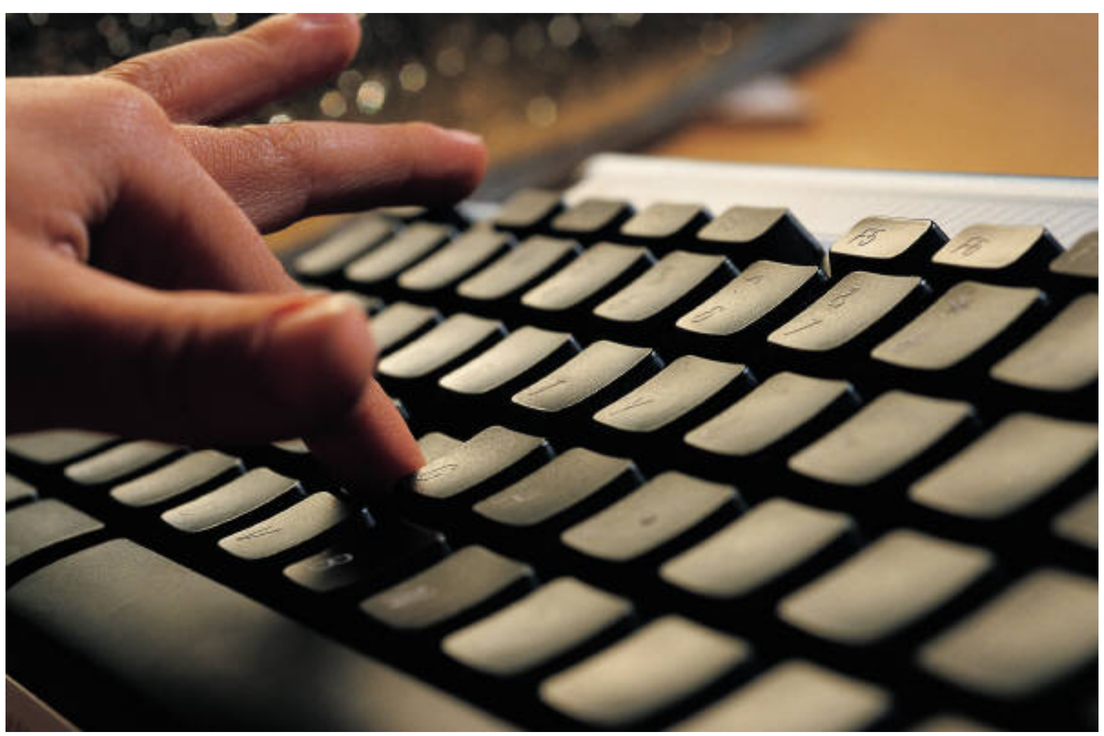
Contrast

- 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)

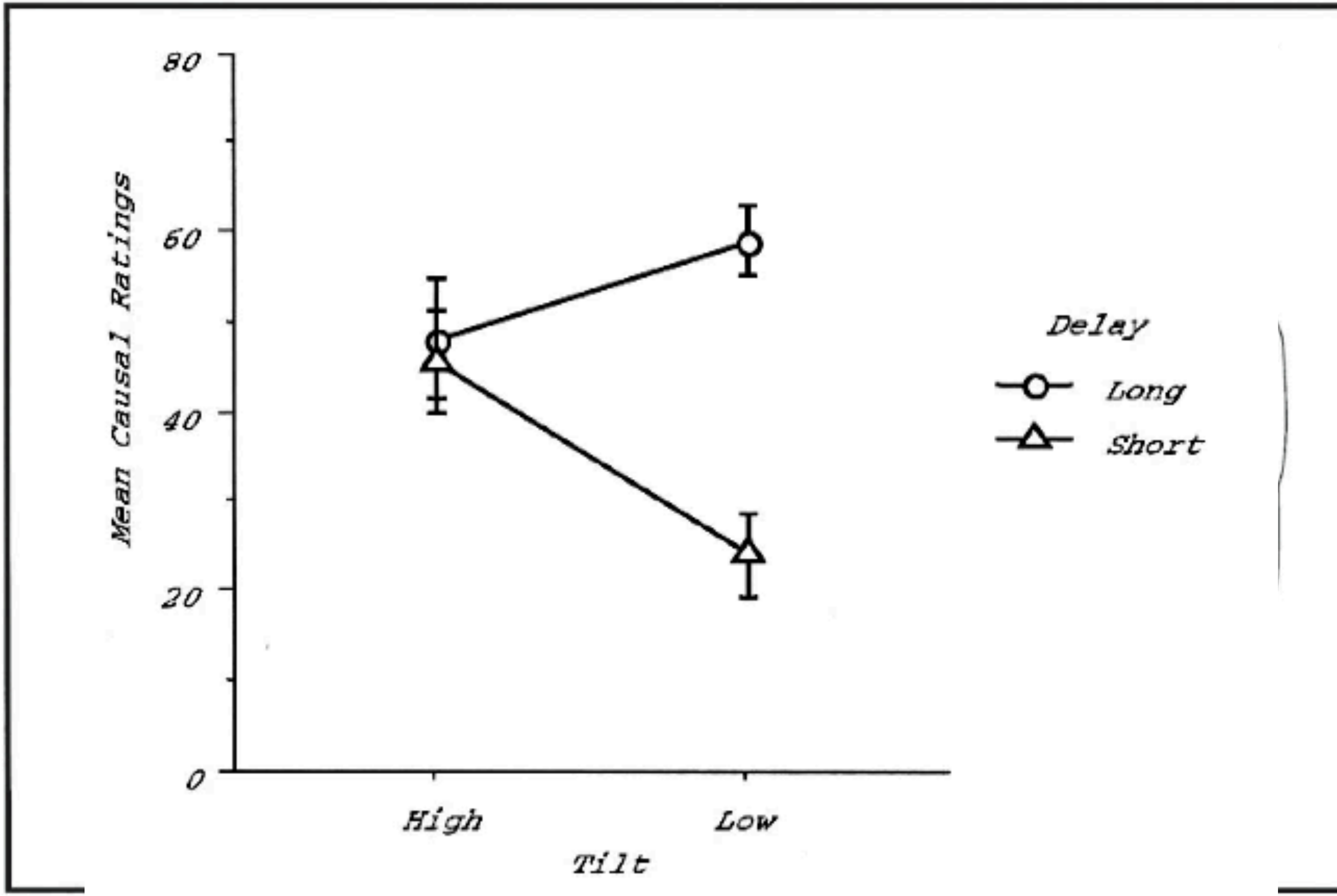


Figure 3. Mean causal ratings from Experiment 1.

Figur ... er on the right.

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.

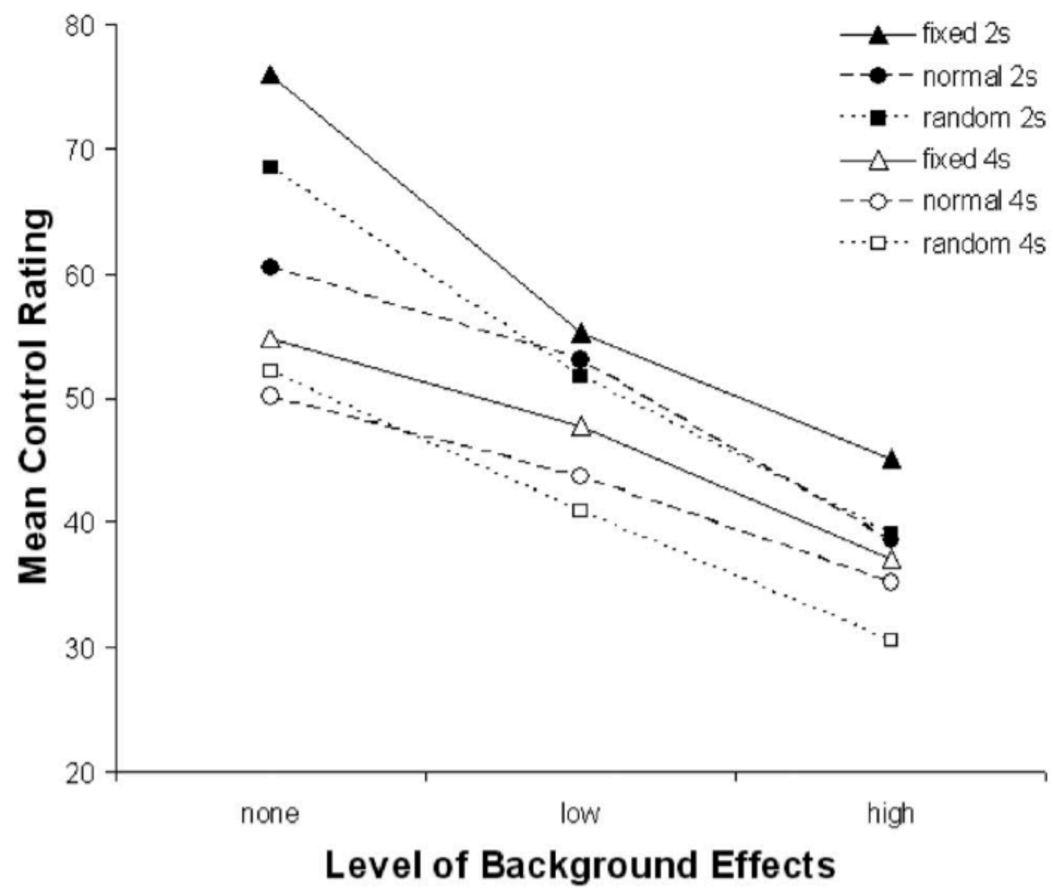
- Time used to learn causality (but can incorrectly override covariation)
- More consistent vs variable timing

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

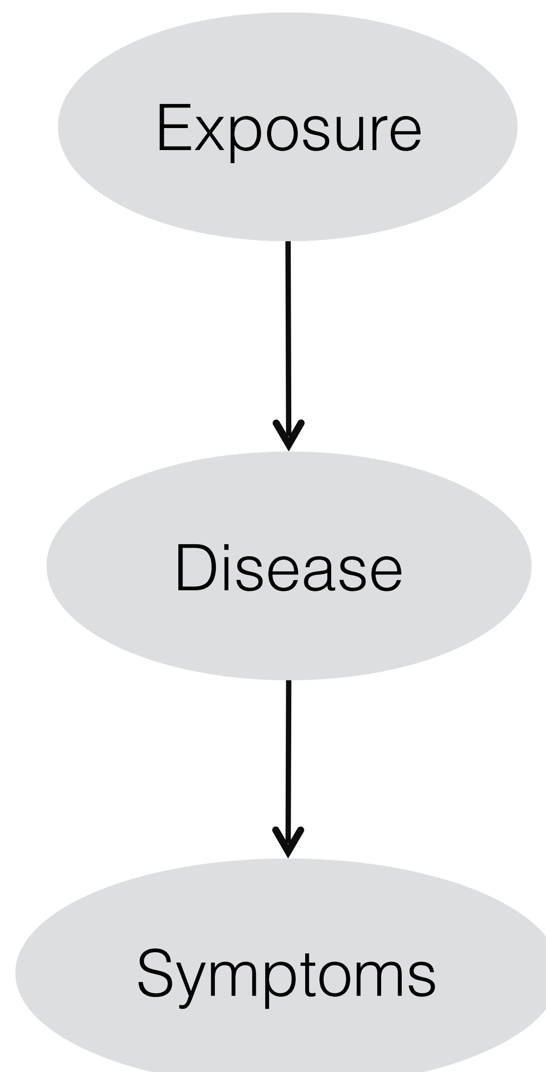
David A. Lagnado* and Maarten Speekenbrink

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.

Time in graphical models



First-order Markov process

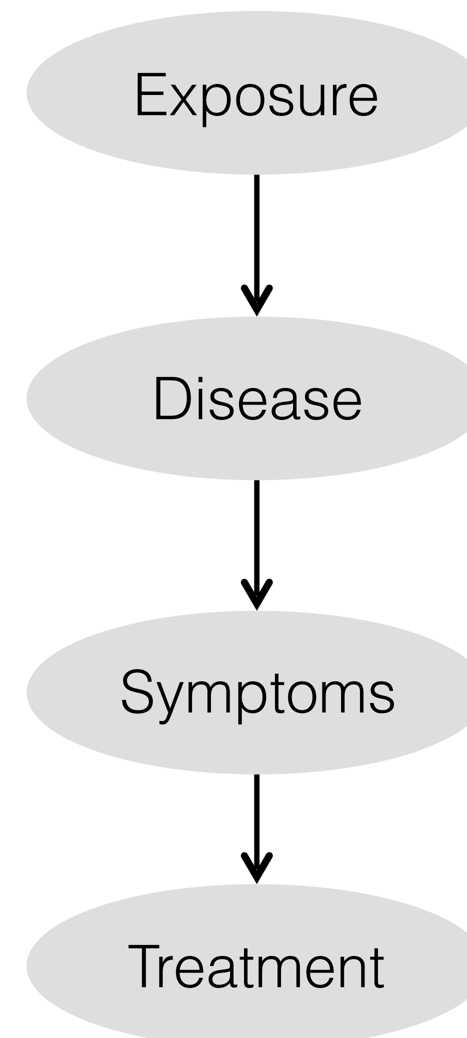
- Edges only between i and i+1

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

- 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

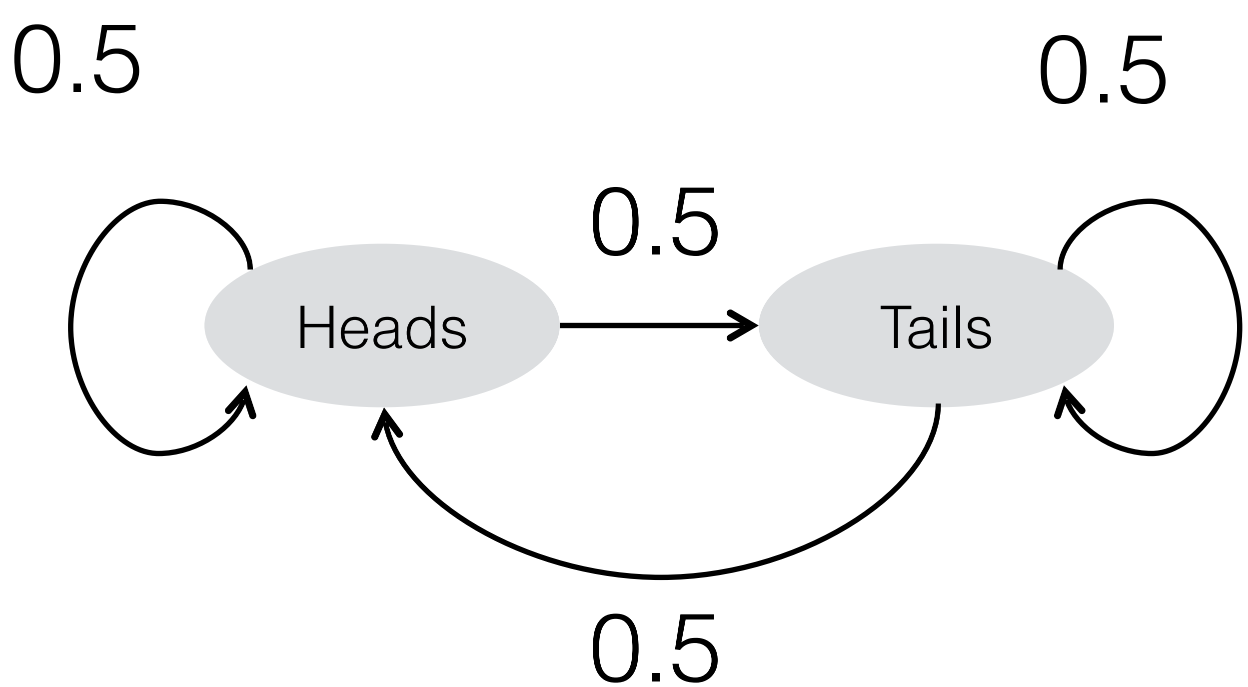


- Path to current state doesn't matter
- Once current state is known, information on prior states not informative

Terminology

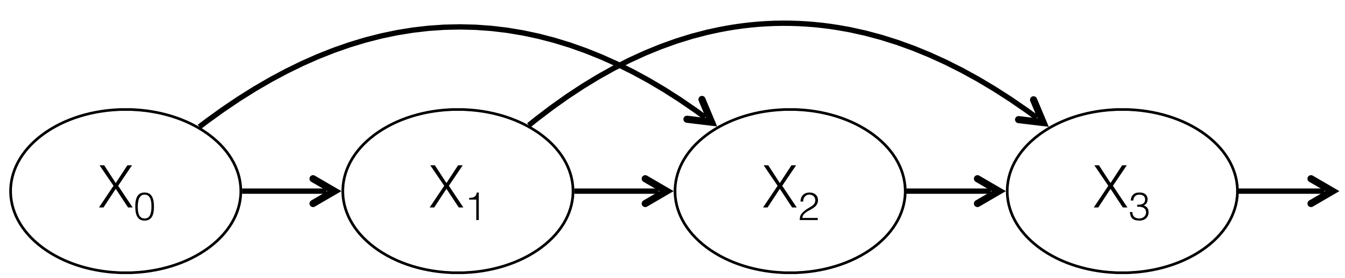
Markov process: stochastic process where transition probabilities dependent only on current state

Each node is a state (collection of variable assignments)



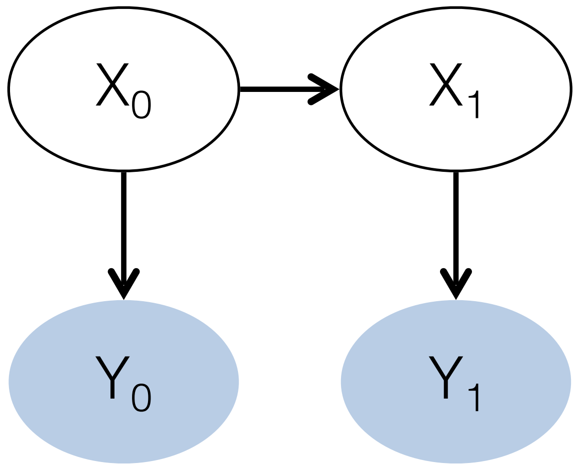
General

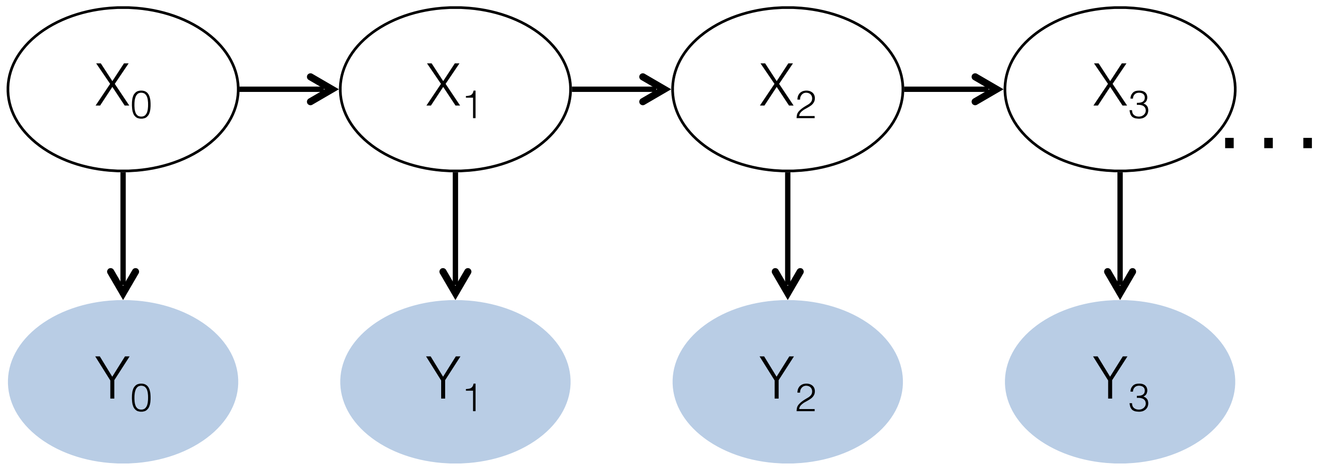
- First order: t depends on t-1
- Second order: t depends on t-1 and t-2



Hidden Markov Model

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

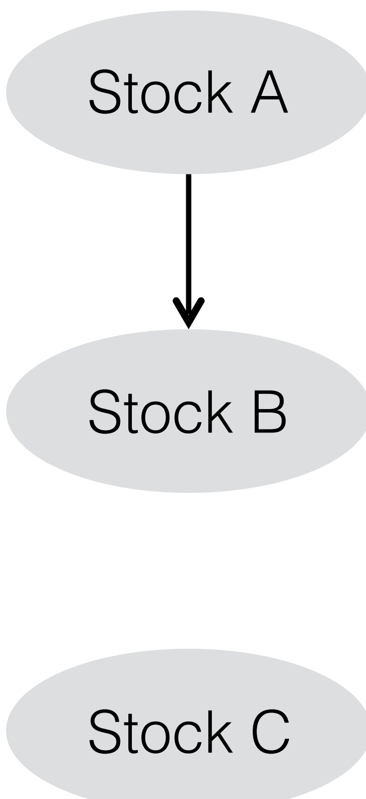




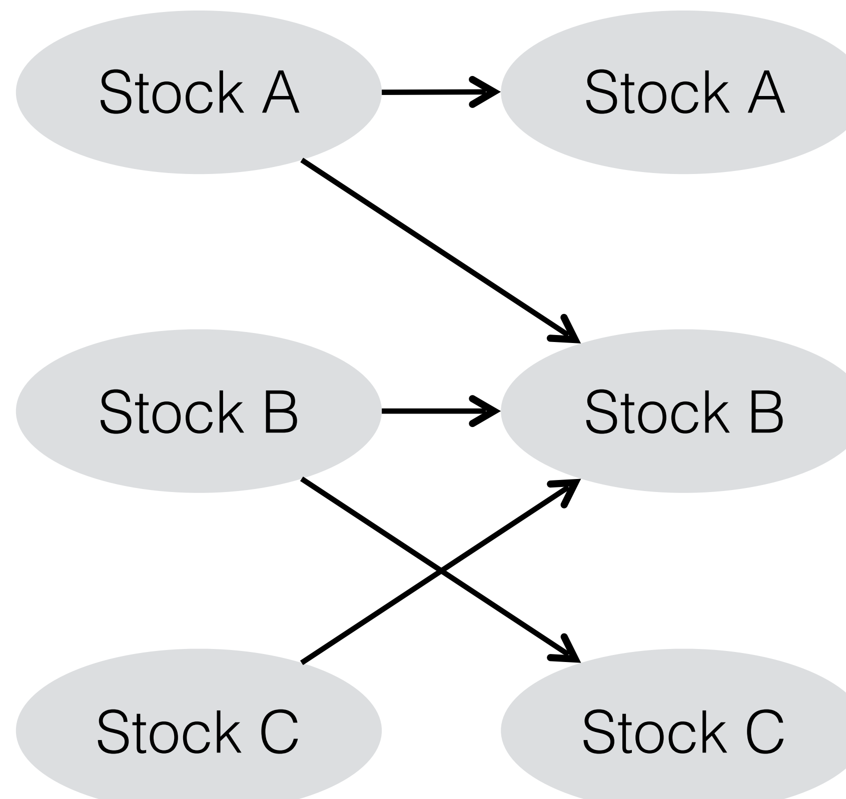
Dynamic Bayesian networks

- “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

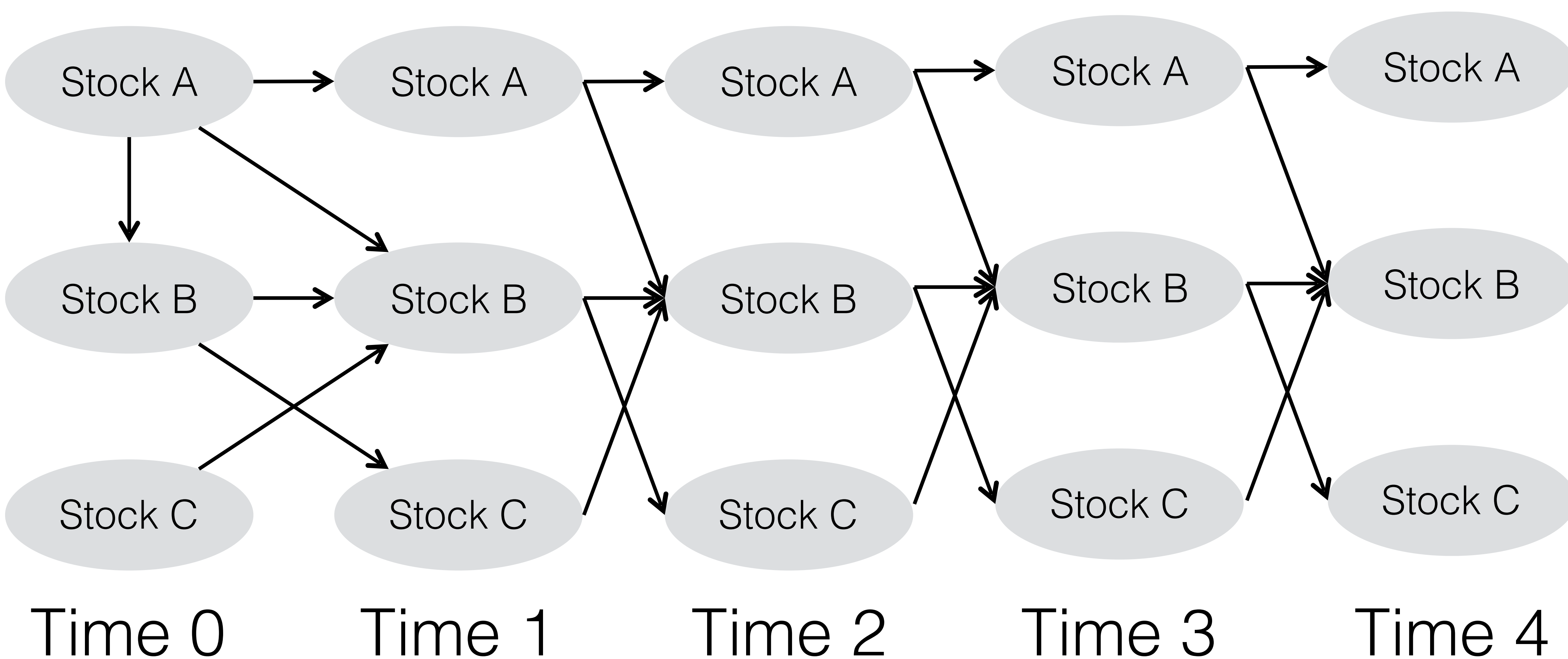


Initial



Transition from t to t+1

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

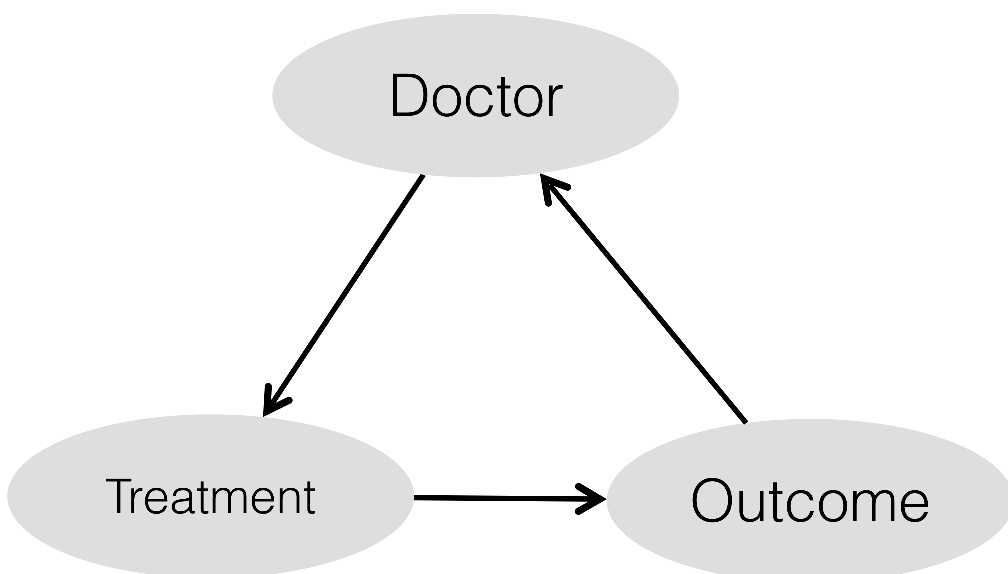
DBNs

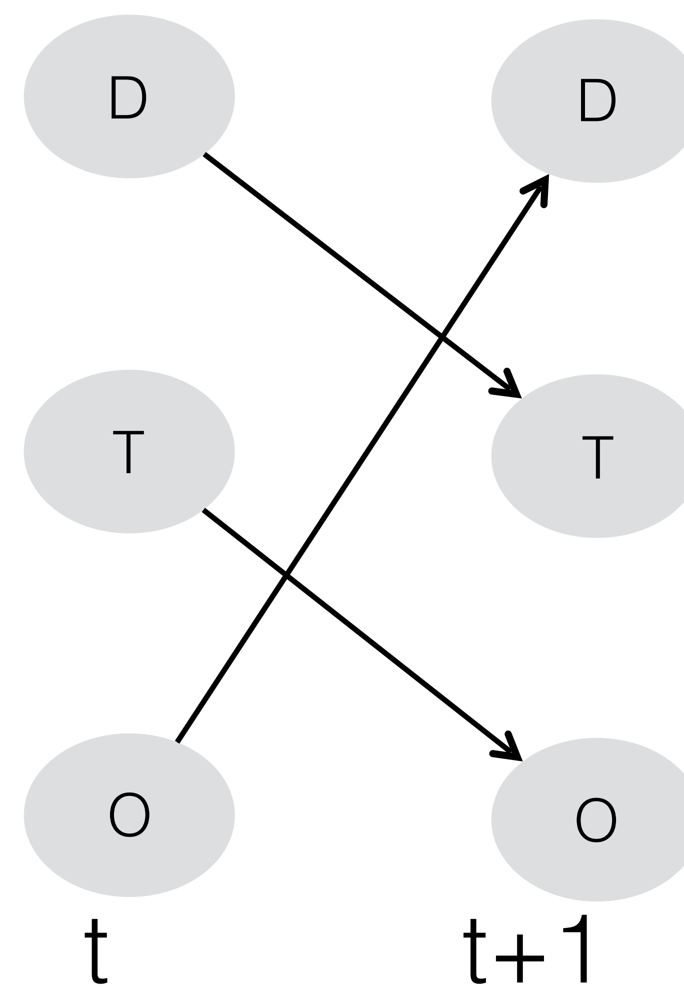
A DBN is defined by:

-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

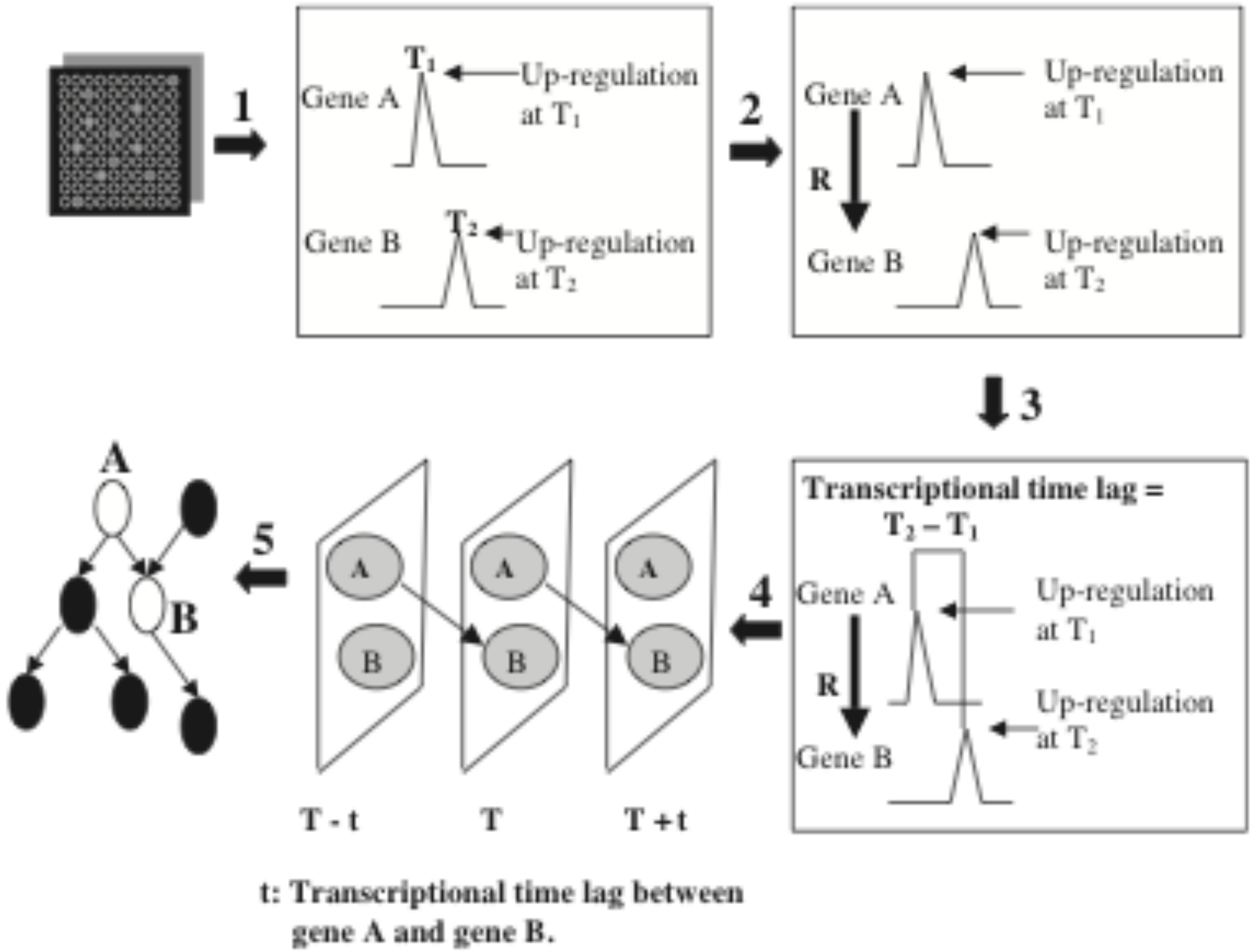
Grzegorczyk, M., & Husmeier, D. (2009). Non-stationary continuous dynamic bayesian networks. *Advances in Neural Information Processing Systems (NIPS)*, *22*, 682-690.

Some software

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

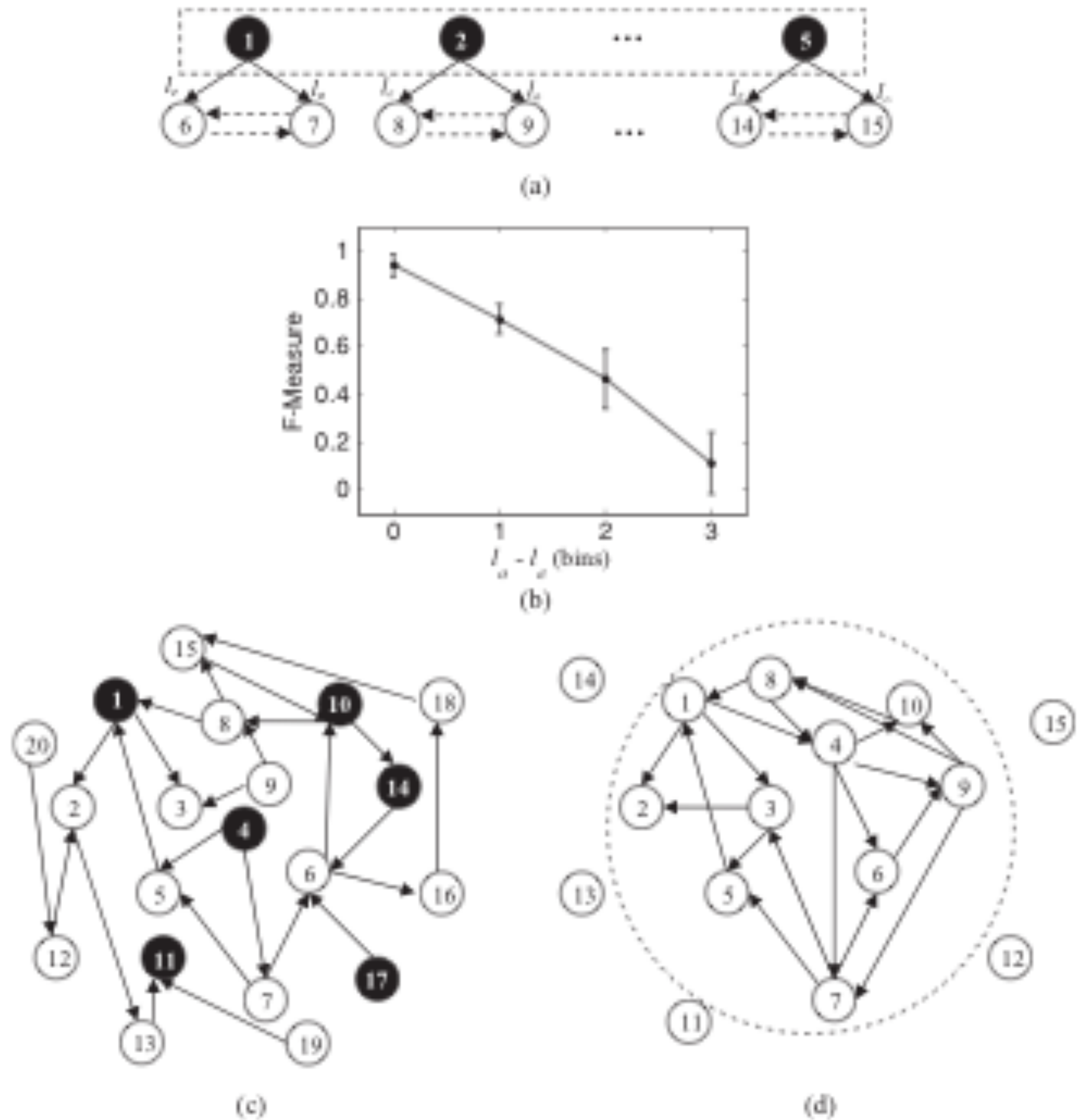
More: http://www.cs.ubc.ca/~murphyk/Software/bnsoft.html

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

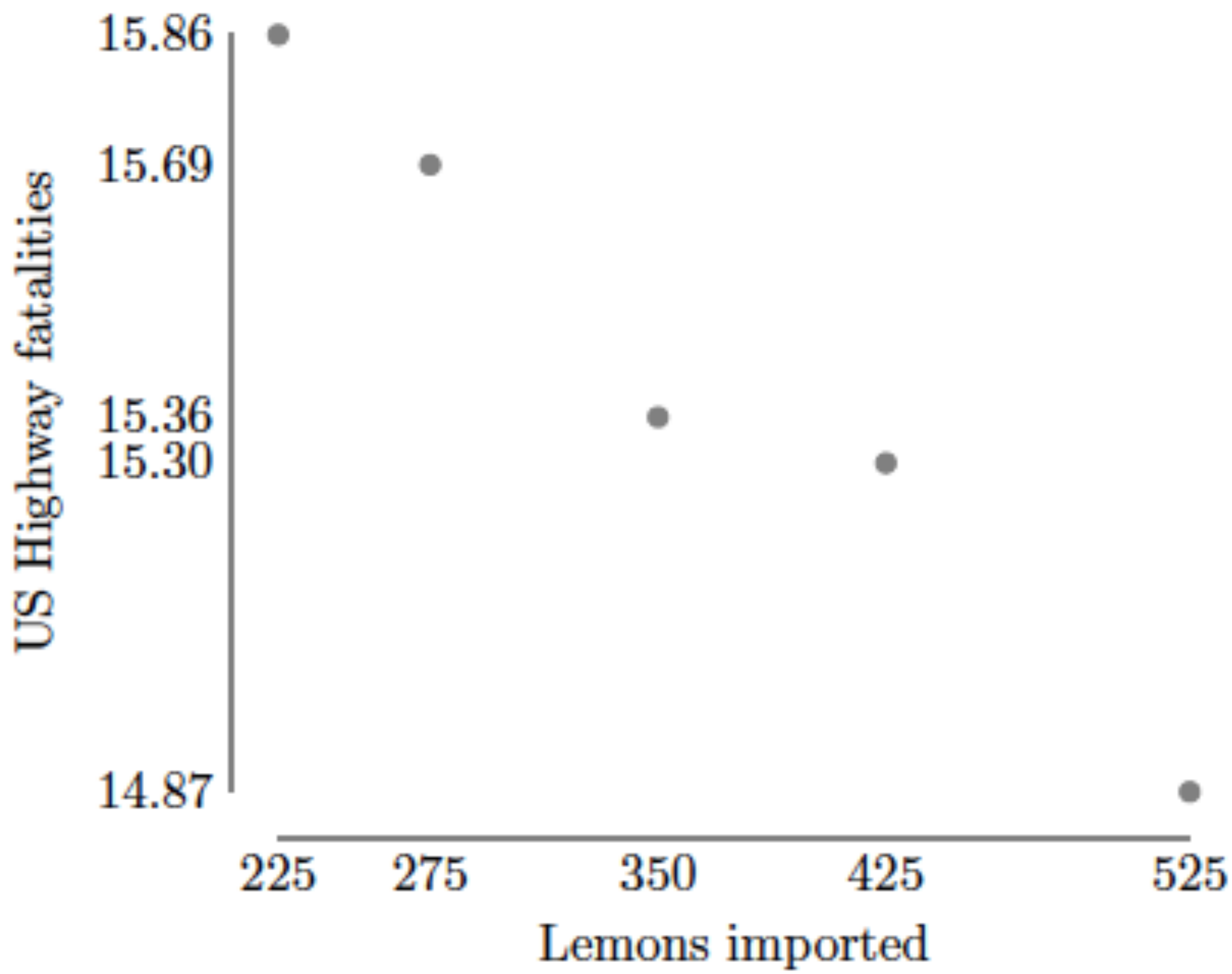


Eldawlatly, S., Zhou, Y., Jin, R., & Oweiss, K. G. (2010). On the use of dynamic bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural Computation*, *22*(1), 158-189

But…

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

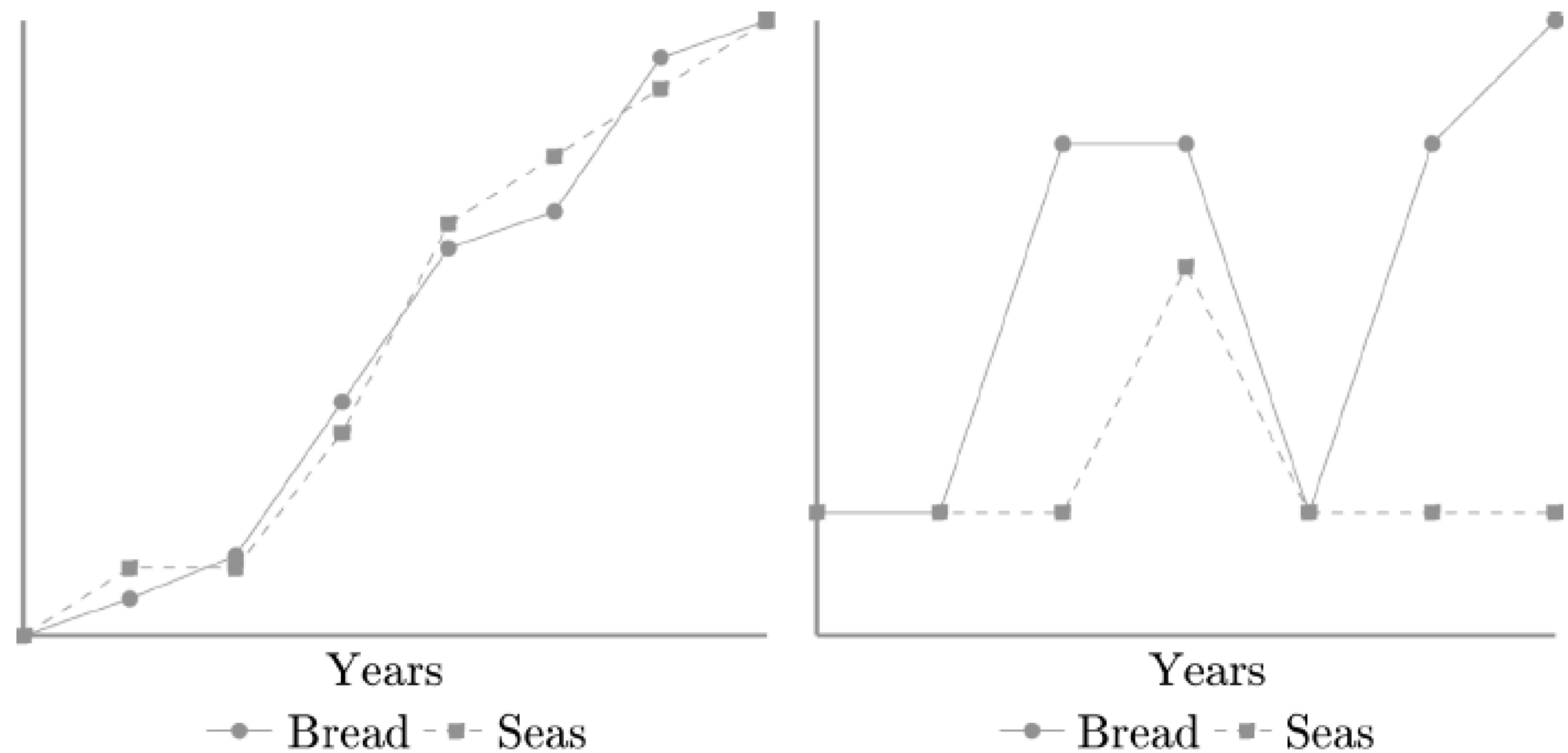
Nonstationarity



- 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)



(a) Raw data *(b) Differenced data*

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:
Kleinberg, S (2013). *Causality, Probability and Time*, Cambridge University Press.

What is a causal relationship?

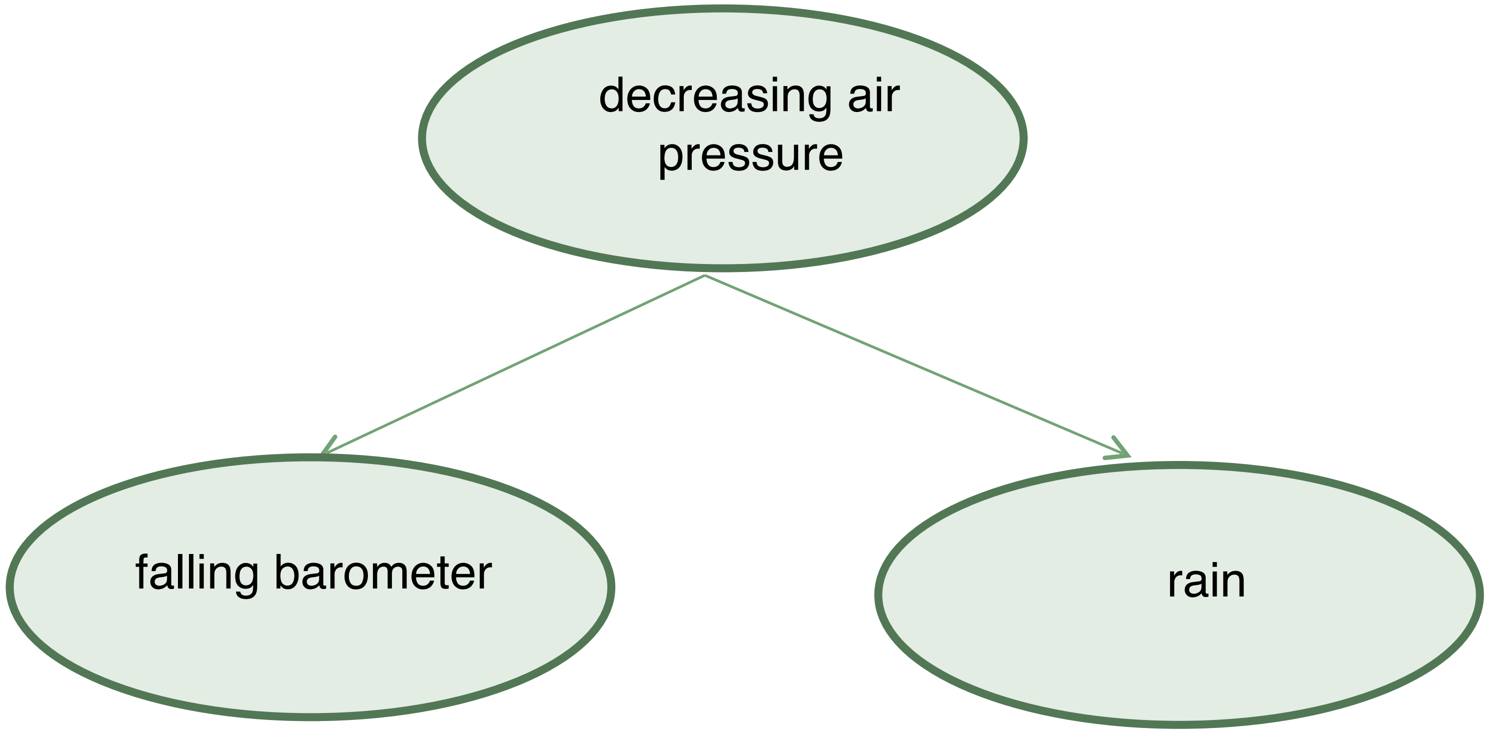
smoking and hypertension until a particular genetic mutation occurs causes CHF in 1-2 years with probability p

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

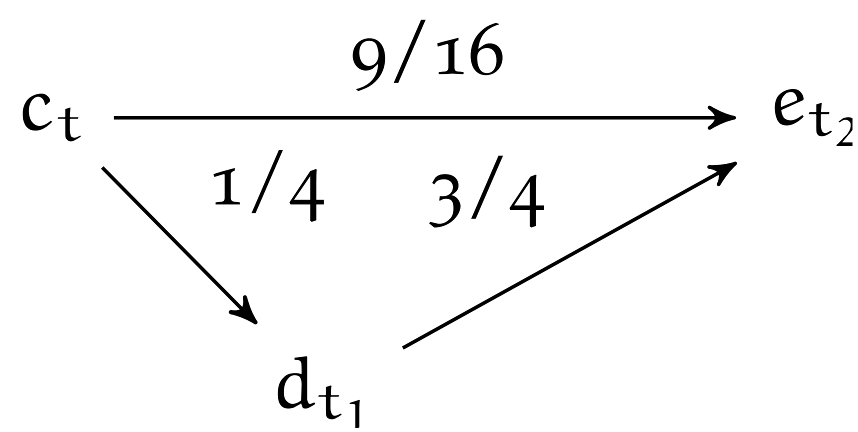
*made up relationship + timing

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

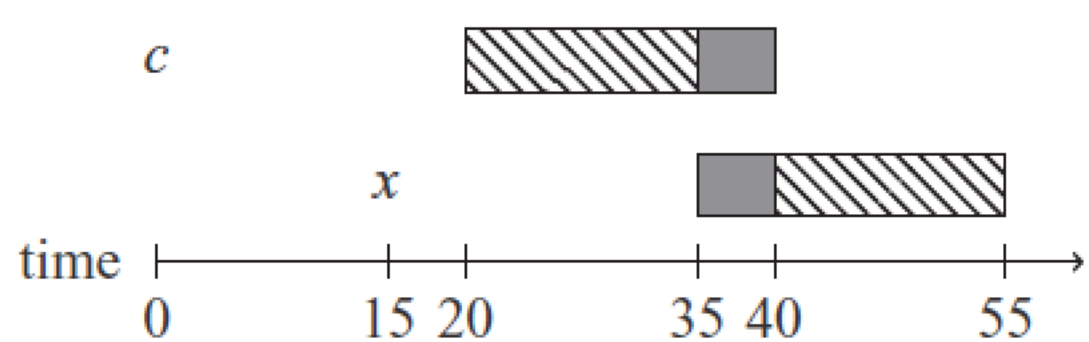
Causal significance

- Potential causes raise probability P(e|c)>P(e) or change expected value E[e|c]≠E[e] PCTLc: Kleinberg 2011

- Assess average difference cause makes to probability of effect

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

- Note time:

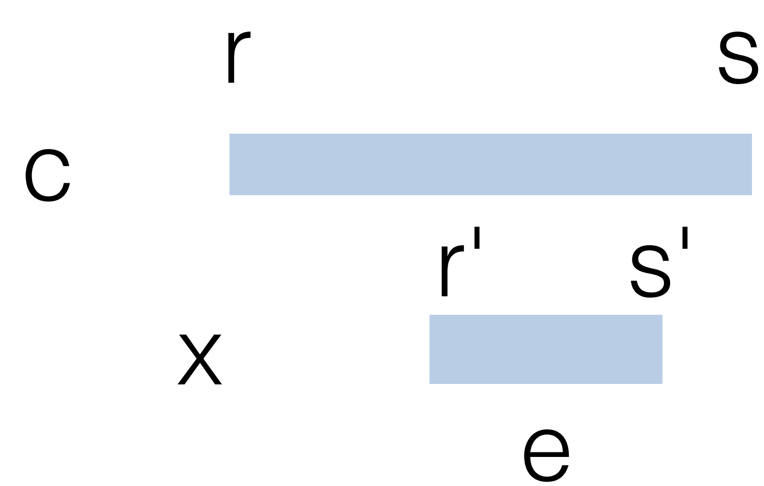


ε_{avg} and time

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

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

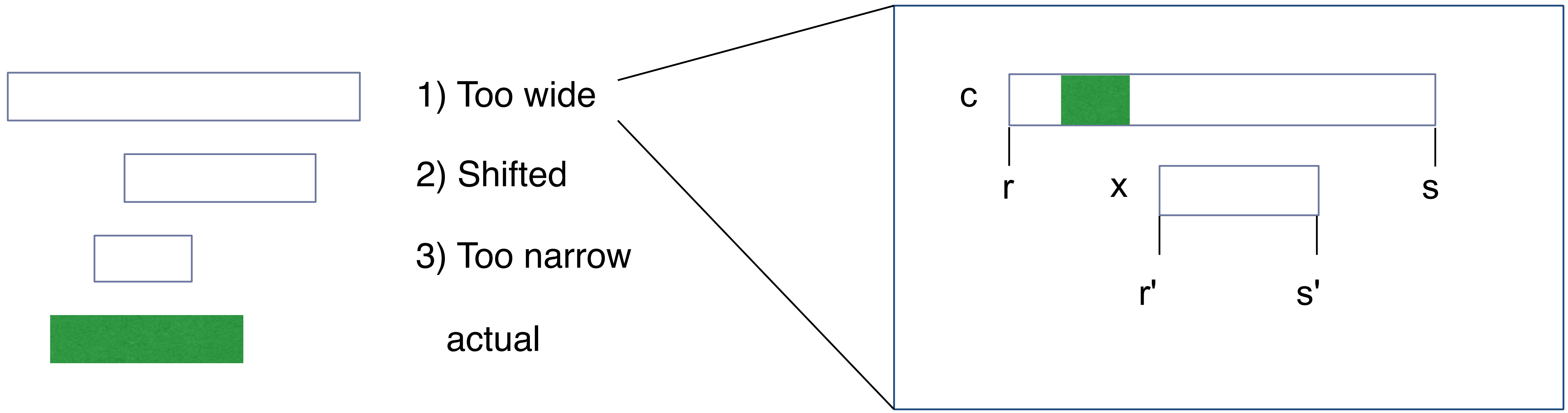
P(e|c^x) means c,x both occur and e occurs in overlapping time window



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



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

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

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

- Key assumptions
  - Stationarity, no latent confounders
- Main advantages
  - Exact inference, time window (vs. lag), complex relationships
- Complexity: $O(N^3T)$

More on ε_{avg}: set X

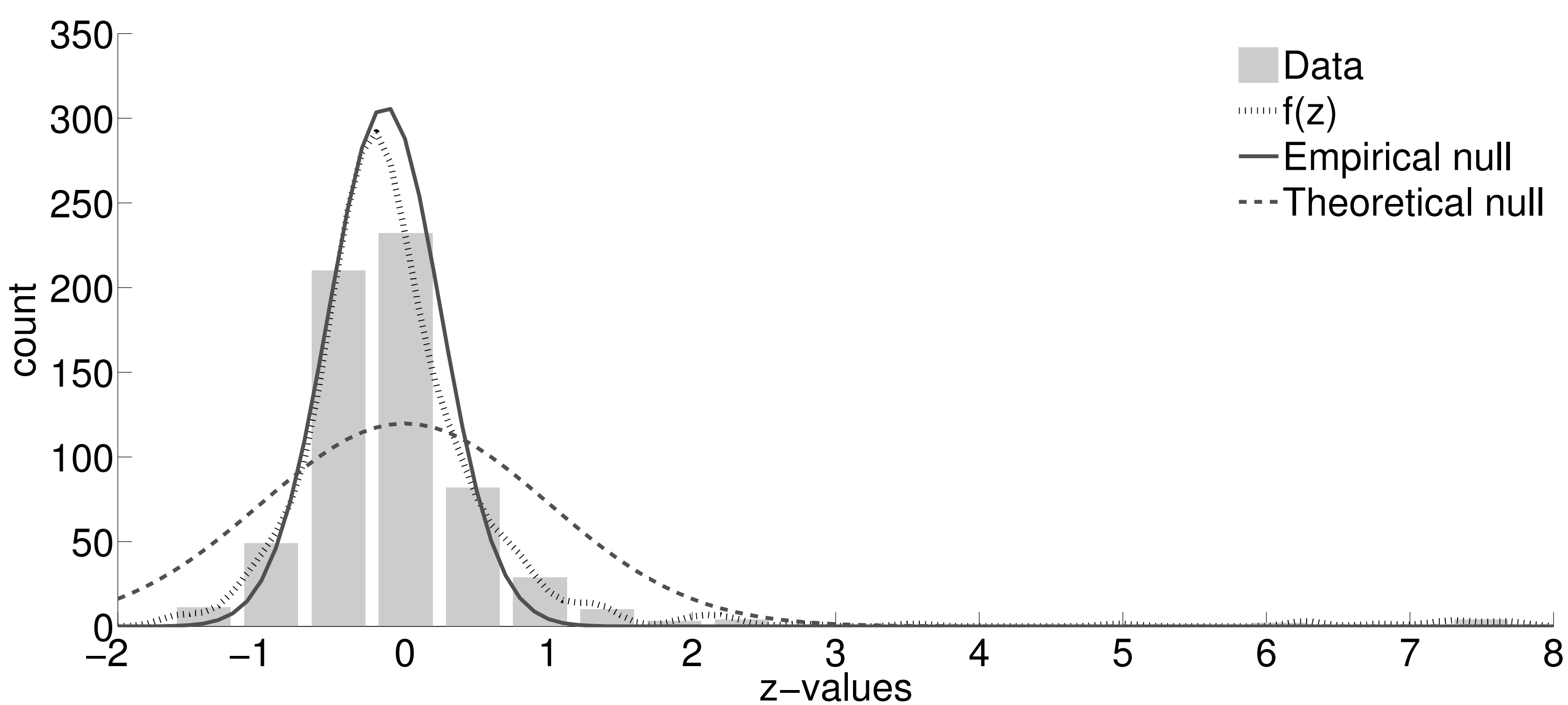
- What is compared against?
  - Probability raisers at any time before e
  - Could be causes/effects of c
- Why can we do this? Two cases:
  - Independent
  - Negatively correlated

What does ε_{avg} mean?

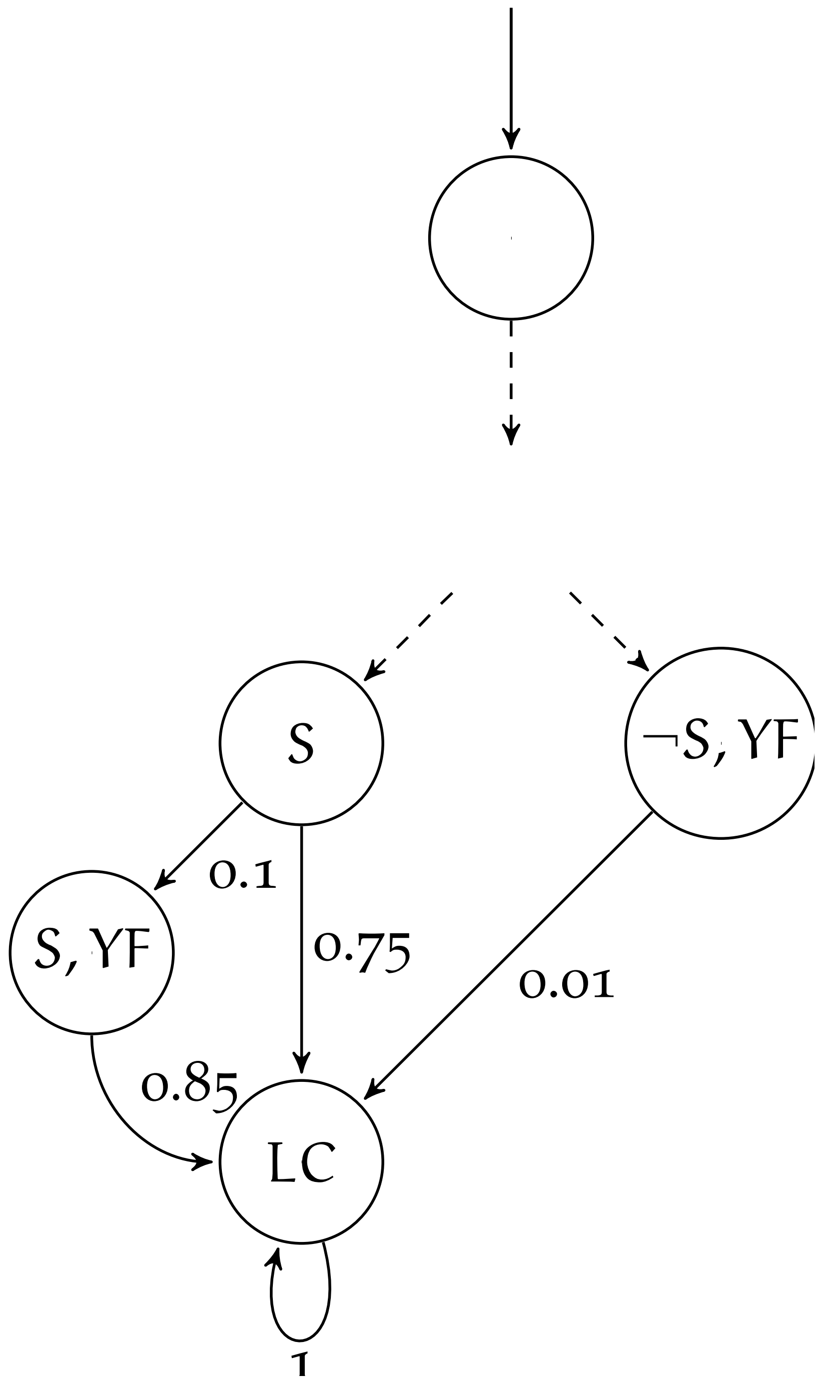
- Positive: c has positive influence on e’s probability
- Negative: probability of e greater after c’s absence (c is possible negative cause of e)
- Zero: +/- may cancel, or no influence
- Small value: maybe artifact, maybe weak real cause

Properties of ε_{avg}

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



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

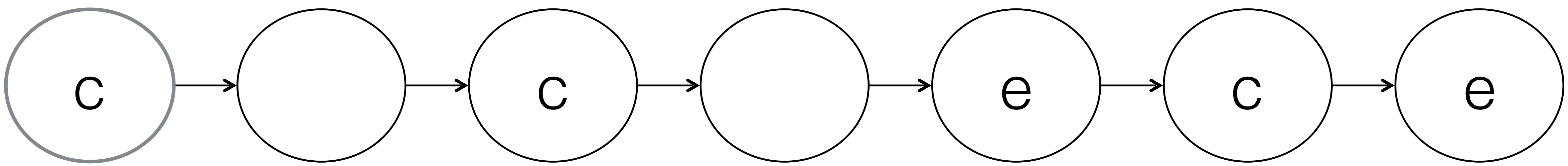
- c is a **just so** or **ε-significant** cause of e if it is a non-ε-insignificant potential cause

- Note that 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: $(s \wedge h) U g$
  - Calculate conditional probability across sequences using frequencies

Probability of c leads-to e in 1-2 time units



False **True** **True**

- Number of times satisfying c: 3
- Number of times satisfying leads-to formula: 2
- Probability = 2/3

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

Satisfied by trace if
p≤ 2/3

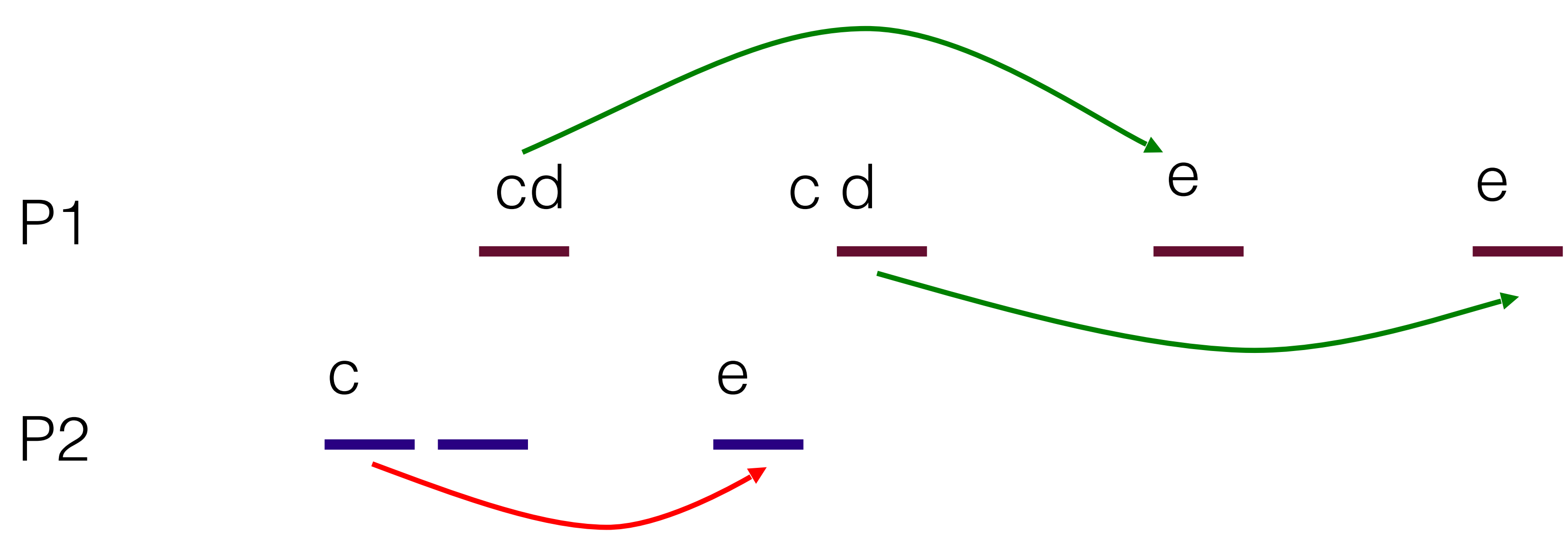
Another example

- To determine significance, need to calculate probabilities

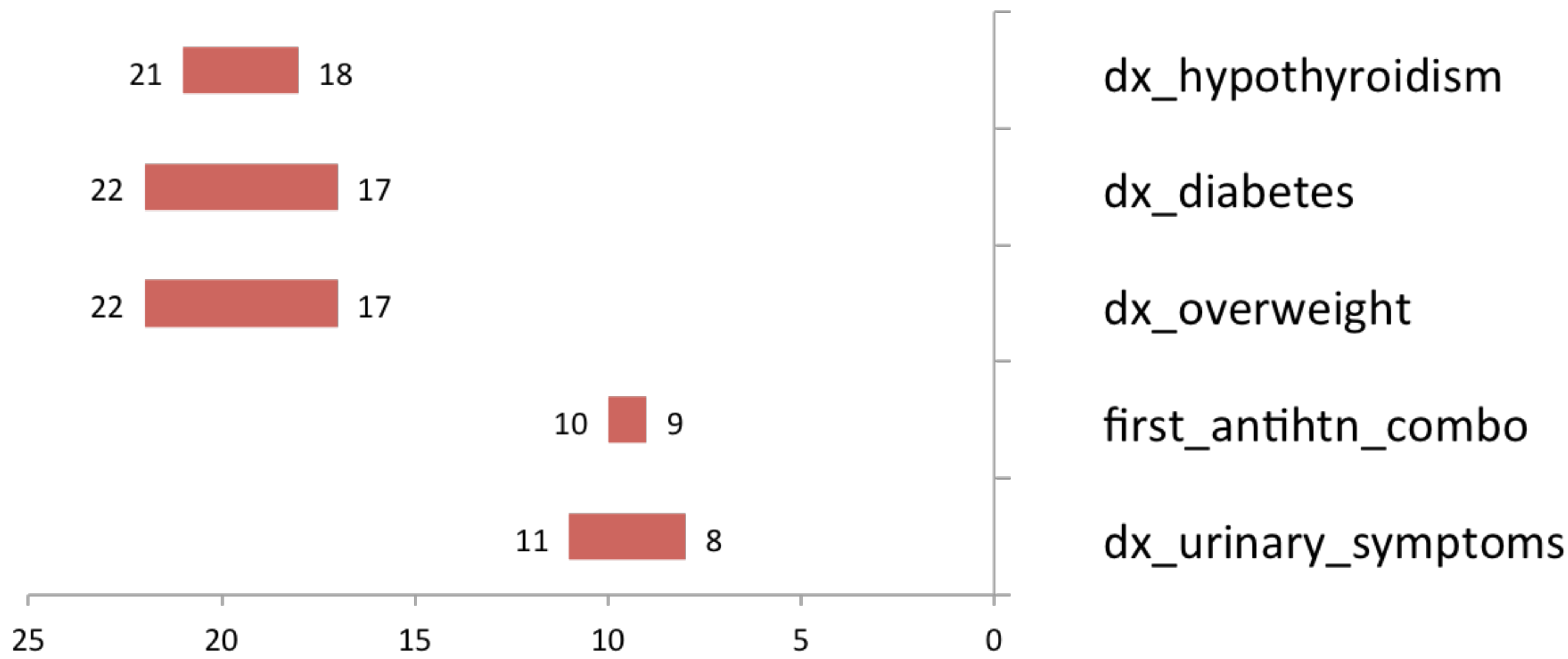
$$f = c \wedge d \overset{\geq s, \leq t}{\leadsto} e$$

- Count frequency, sum over all traces

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



Congestive heart failure



Months before CHF diagnosis

Kleinberg S, Elhadad N (2013) Lessons Learned in Replicating Data-Driven Experiments in Multiple Medical Systems and Patient Populations. In: AMIA Annual Symposium.

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

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

Mining ... ly of

Andrew St ... akarian*

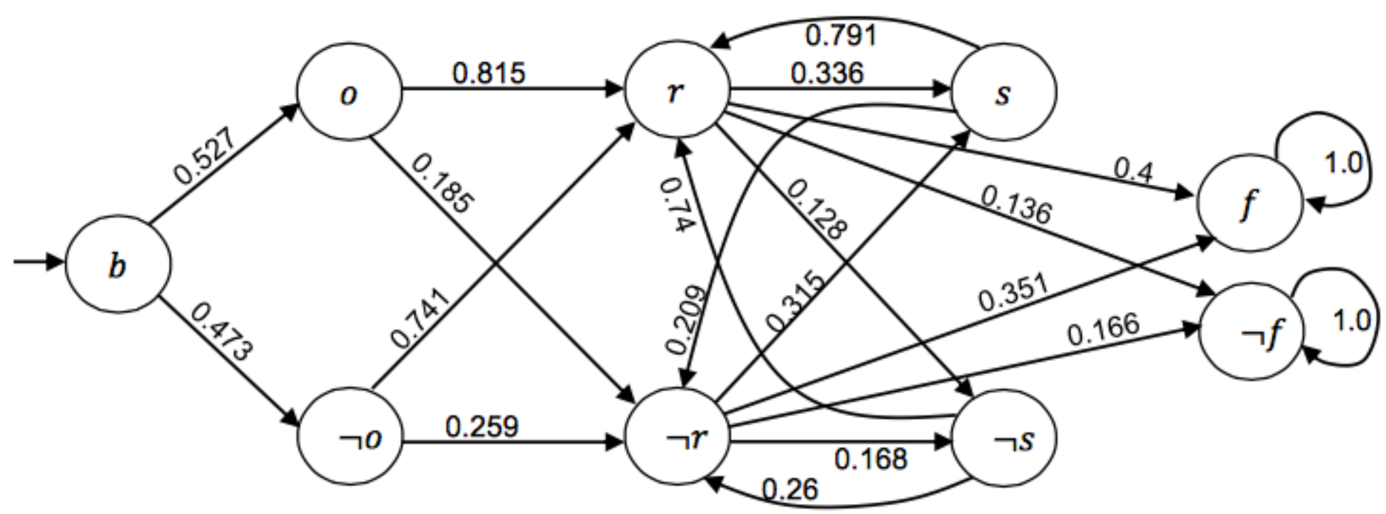


Fig. 2: Probabilistic Kripke Structure for CSN

ABSTRAC

The Islamic St
insurgent grou
prominence wh
paper, we pres
group using a da
tary activity sur
it (including Iraq
We combine idea
soning to mine f
evidence of caus
ISIS vehicle-bour
activity in Syria
air strikes, and IS
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.

ic propo-
mparing
o consid-
measure
in [10]) –
age increase
considered
pproach, we
e following:

tions in Iraq
ndicative of
e (VBIED)
k (p = 1.0,

	armedAtkSpike($Syria, \sigma$)			

erations in Tikrit

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($Iraq, \sigma$)$\wedge$ indirFireSpike($Iraq, 2\sigma$)	0.92	1.00	0.20
6.	armedAtkSpike($Iraq, \sigma$)$\wedge$ indirFireSpike($Iraq, 2\sigma$)$\wedge$ VBIED($Baghdad$)	0.92	1.00	0.20

(paper that was for today)

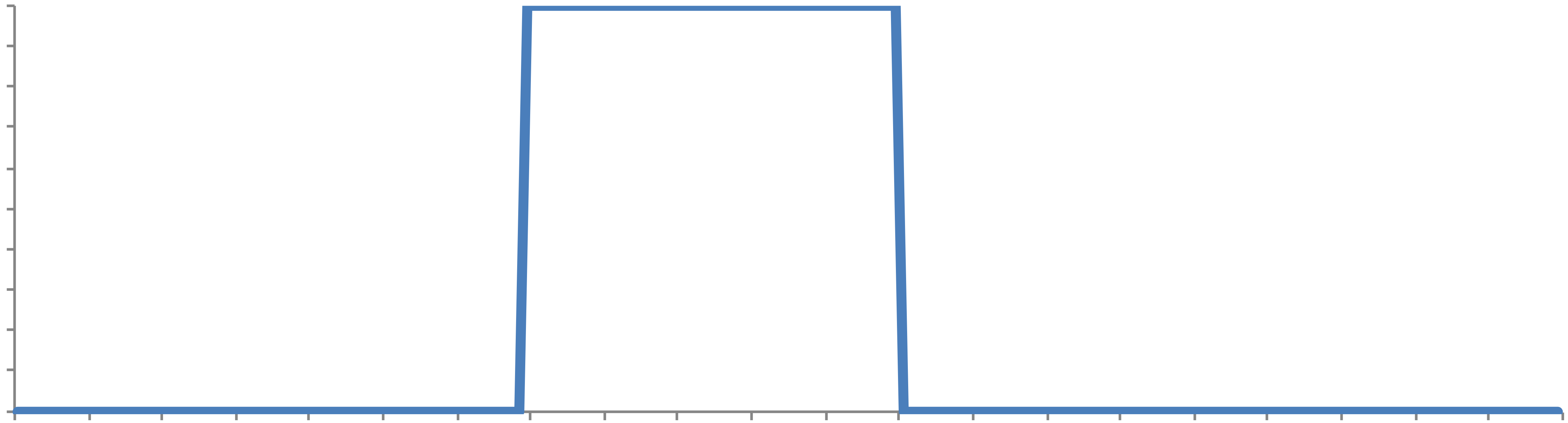
- Chronic diseases are managed primarily by individuals — adjusting insulin dosing, making and logging food choices
- Hospital data is episodic
    - Is patient on medication between visits? Are they following dietary recommendations?
- Outpatient data brings more uncertainty

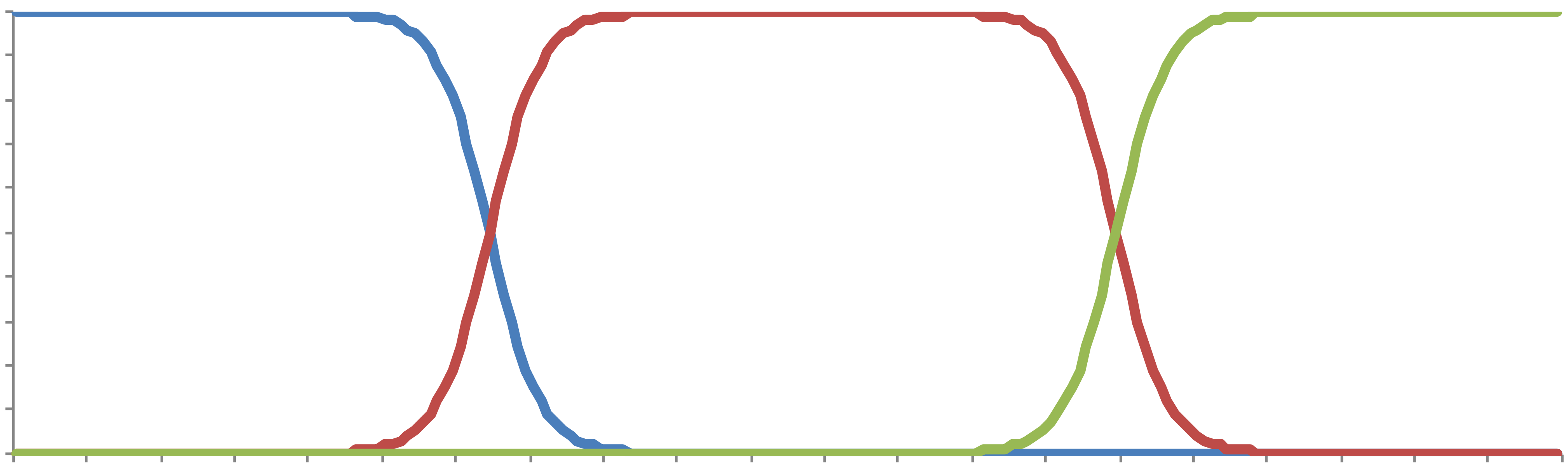
Partial solution: probabilistic observations

- Issues
  - Error in measurements
  - Delay
  - Inconsistent timescales

- Strategy
  - Instead of true/false, use probability of event at each time

Discretization





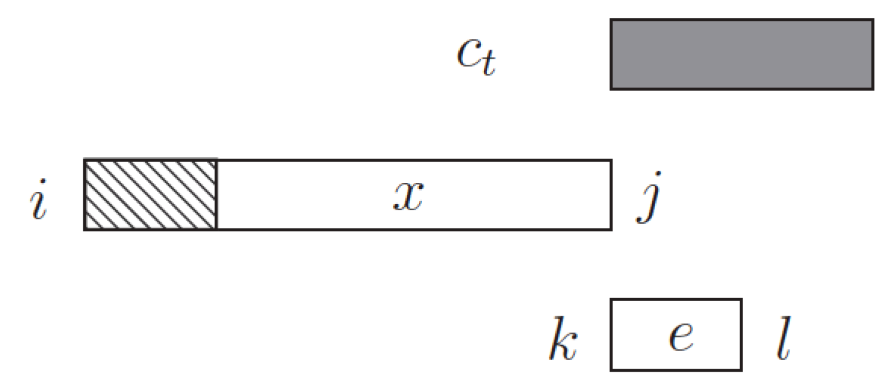
Adding uncertainty to causal inference

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

$$P(e|c,x) = \frac{\sum_t P(e,c,x)}{\sum_t P(c,x)}$$

detail with time windows

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



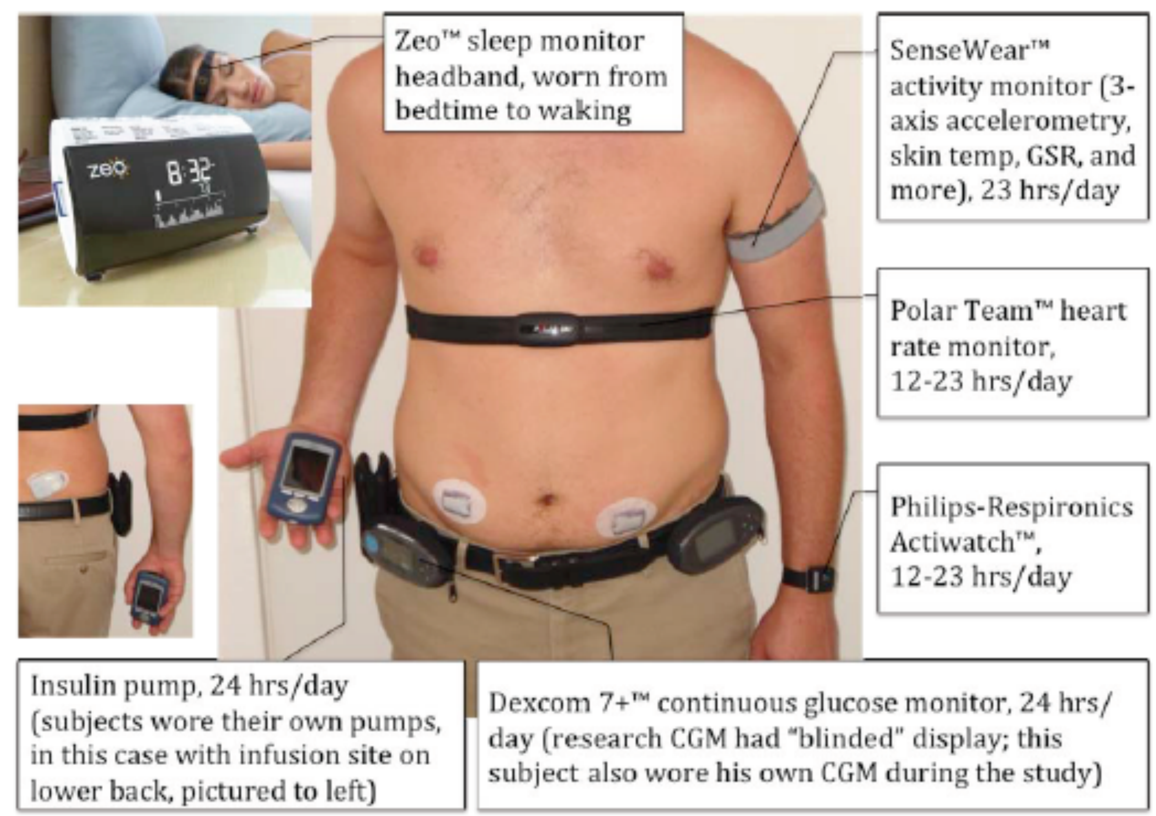
$$P(e|\neg c \wedge x) = \frac{\sum_{t \in T} P(\neg c_{g \wedge \ldots h}) P(x_t) P(e_{k \vee \ldots l})}{\sum_{t \in T} P(\neg c_{g \wedge \ldots h}) P(x_t)}.$$

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

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)

N. Heintzman and S. Kleinberg. Using Uncertain Data from Body-Worn Sensors to Gain Insight into Type 1 Diabetes. Journal of Biomedical Informatics 2016

Explanation

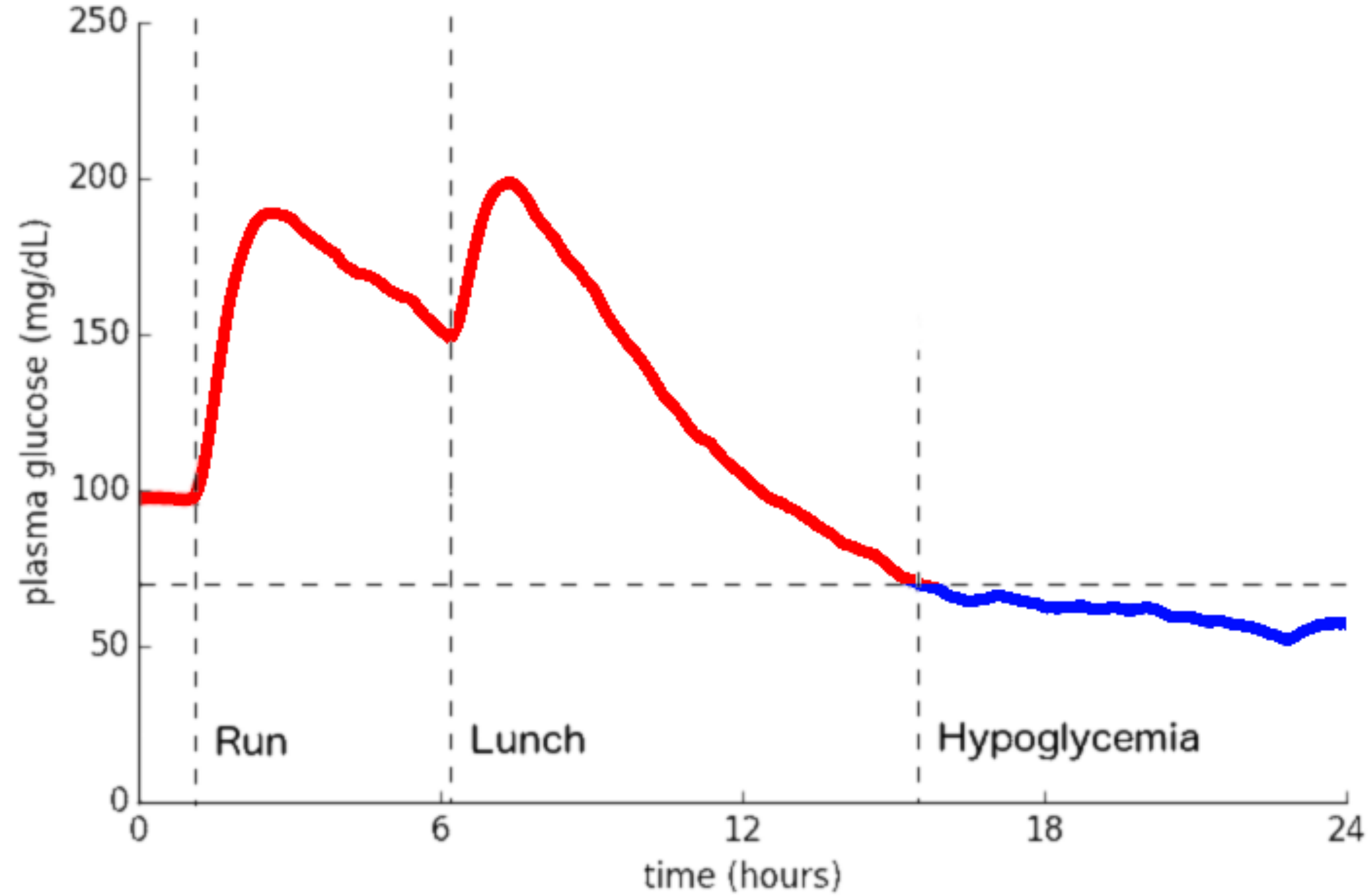
- Where do explanations come from?
  - Need formal representation for algorithms

- Type ≠ Token
  - Cannot assume observations will exactly fit what is known

- Data
  - Frequently missing
  - Time of event ≠ time event is recorded

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?



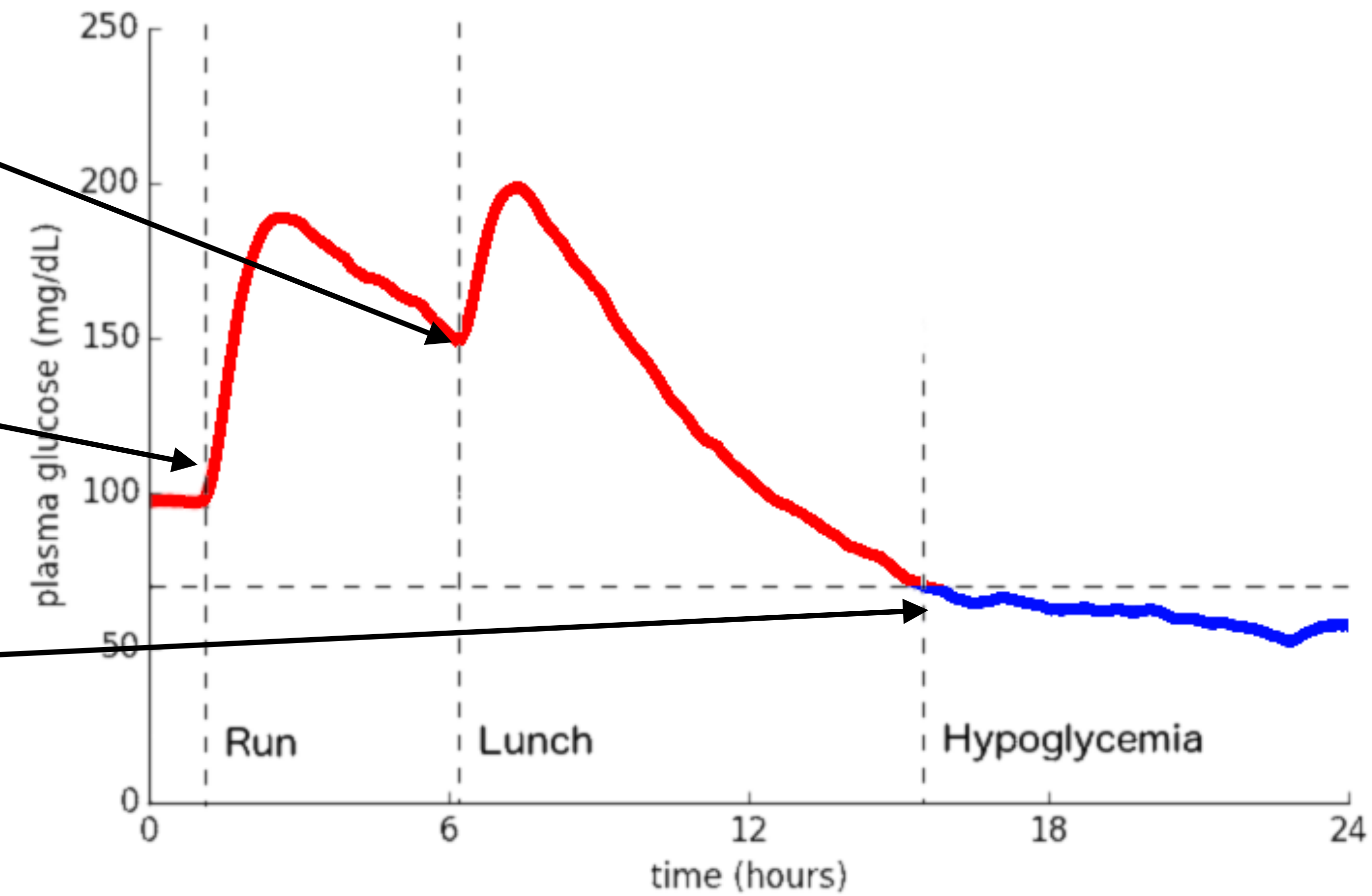
Example

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

C. Merck and S. Kleinberg. Causal explanation under indeterminism: A sampling approach. AAAI, 2016.

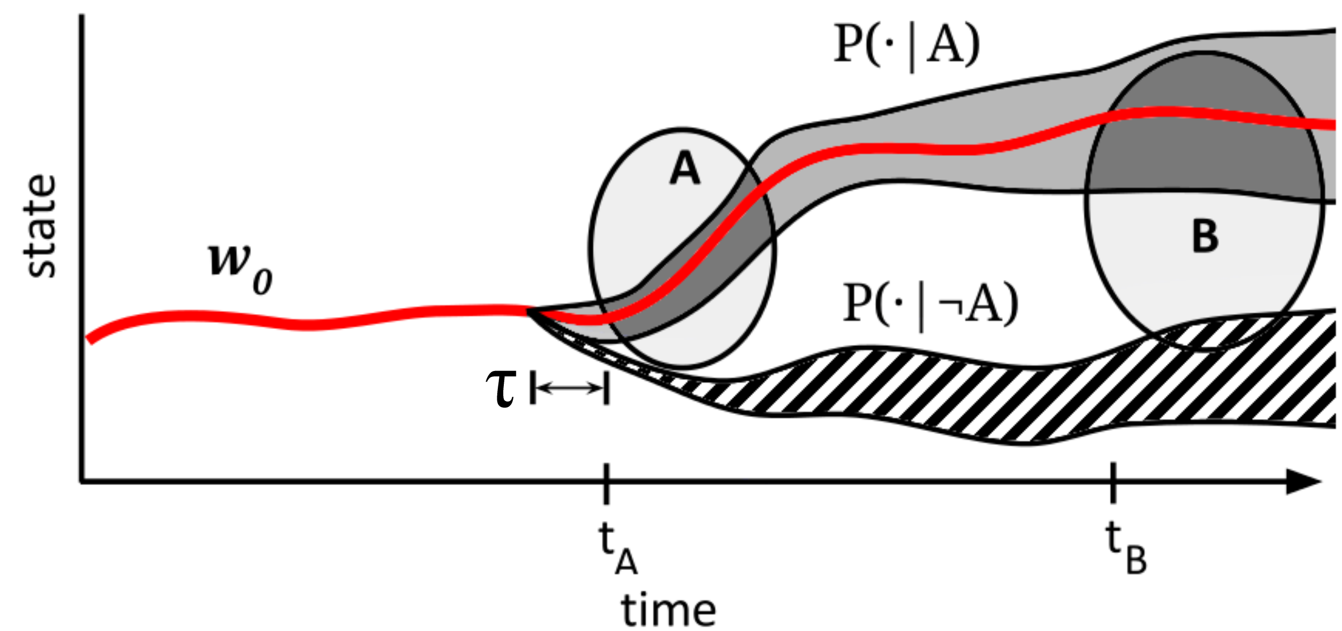
Counterfactual vs. Actual Distributions

- P(•|¬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(•|A) is the *actual distribution* of A

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



C. Merck and S. Kleinberg. Causal explanation under indeterminism: A sampling approach. AAAI, 2016.

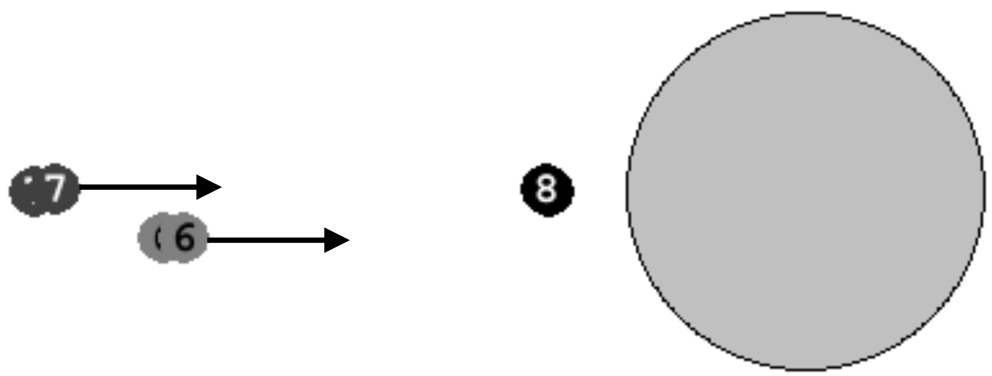
Three Types of Explanation

probability:	B **because of** A iff $P(B \mid A) \gg P(B \mid \neg A)$	B **despite** A iff $P(B \mid A) \ll P(B \mid \neg A)$
timing:	B **hastened by** A iff $E[t_B \mid A] \ll E[t_B \mid \neg A]$	B **delayed by** A iff $E[t_B \mid A] \gg E[t_B \mid \neg A]$
intensity:	B **intensified by** A iff $E[m_B \mid A] \gg E[m_B \mid \neg A]$	B **attenuated by** A iff $E[m_B \mid A] \ll E[m_B \mid \neg A]$

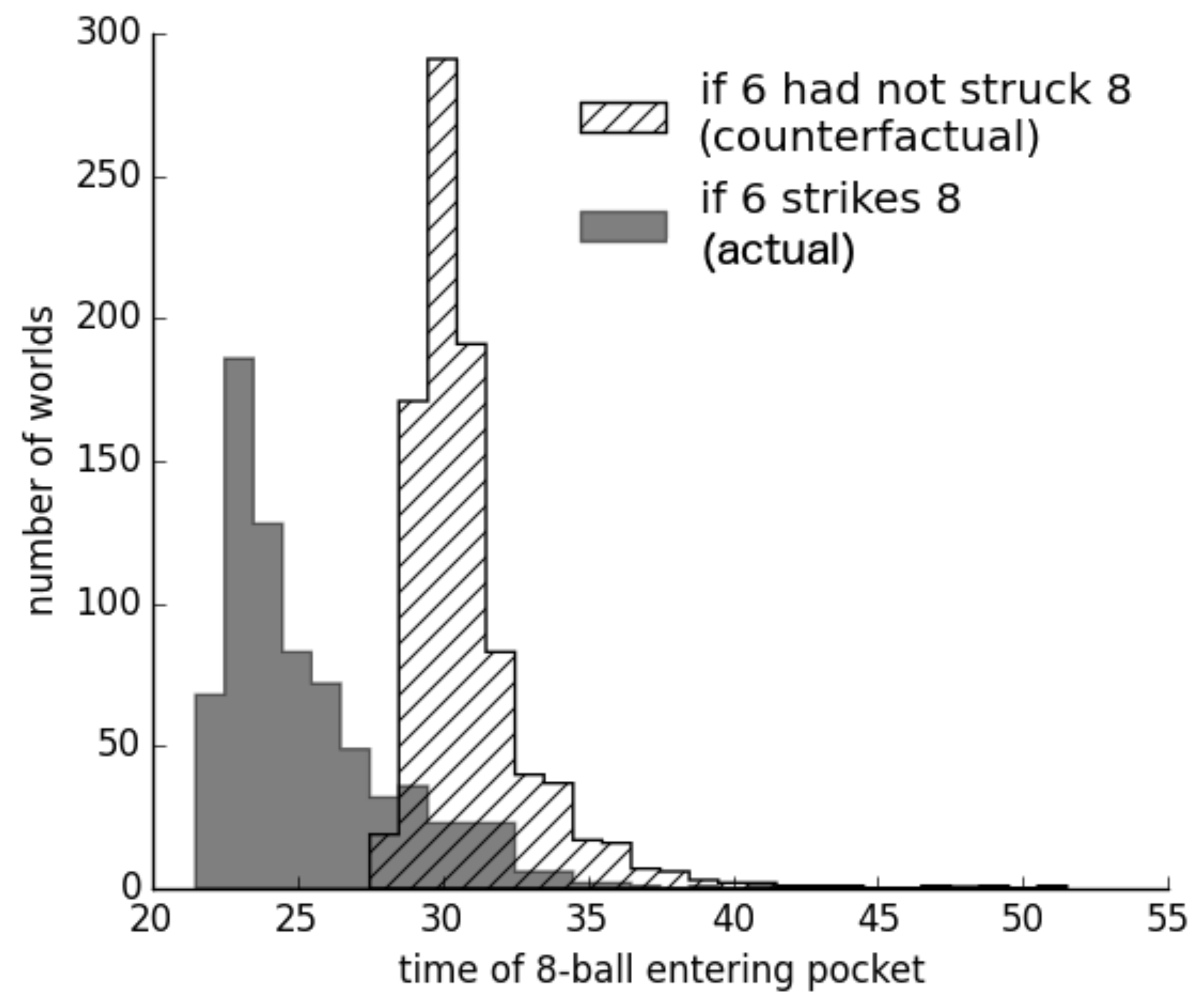
t_B = time of B occurring

m_B = intensity (manner) of B occurring

Hastening

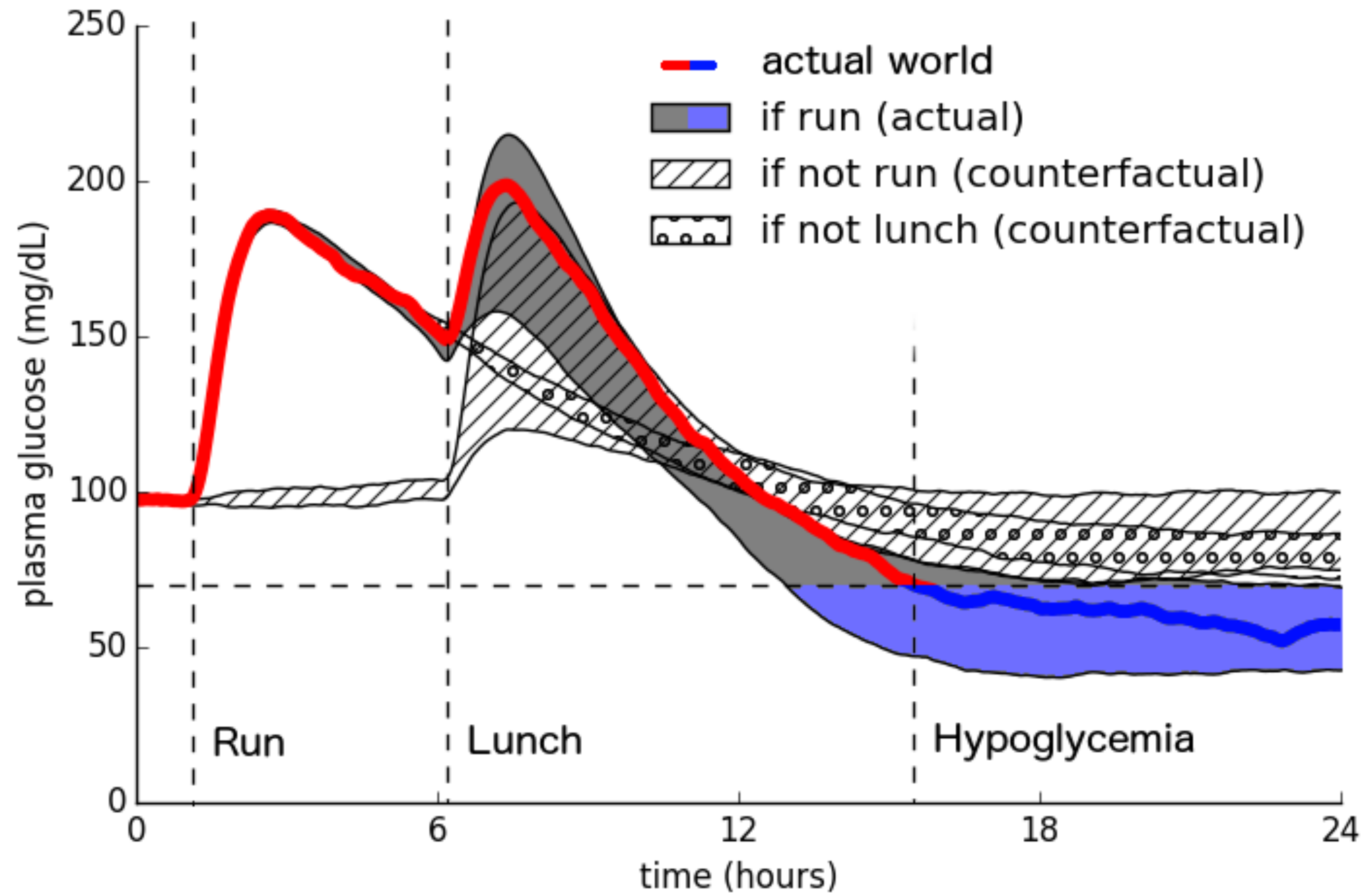


- 6->8, then 8->P
- but 7->8 is a more reliable backup
- probability raising finds
  “8->P despite 6->8”
- but by analyzing timing we find
  “6->8 hastened 8->P”



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:
  $E[\text{ glu} \mid R] - E[\text{ glu} \mid \neg R]$



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
$E[t_B|A] << E[t_B|\neg A]$

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 = time of B occurring

m_B = intensity (manner) of B occurring

For next week

- Discussion paper changed:

  – Discussion paper: B. Elbel, J. Gyamfi, and R. Kersh. Child and adolescent fast-food choice and the influence of calorie labeling: a natural experiment. *International Journal of Obesity*, 35(4):493–500, 2011 [pdf]

- Bring questions for midterm review!!