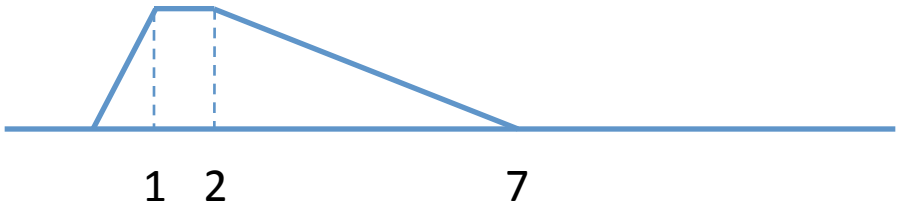
Causal Inference: prediction, explanation, and intervention

Lecture 7: Causality in time series (part II)

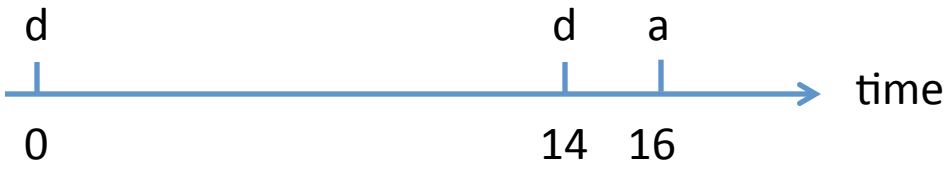
Samantha Kleinberg
samantha.kleinberg@stevens.edu

Example: multiple instances

- Type-level relationships (X), drug (d), adverse event (a) $d \overset{\geq 1, \leq 2}{\rightsquigarrow} a$
- f



- Token observation (V)

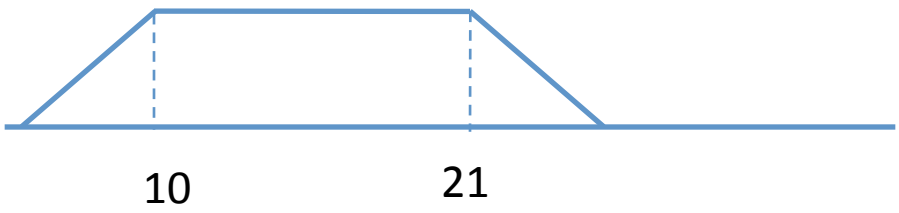


Example: Overdetermination

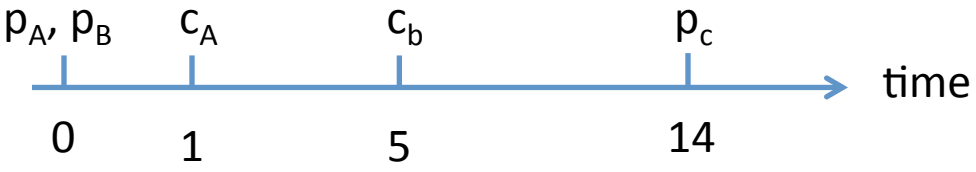
- Type-level relationships (X), contact(c), chickenpox(a)

$$c \overset{\geq 10, \leq 21}{\rightsquigarrow} p$$

- f



- Token observation (V)



Example: Transitivity

- Type-level relationships (X)

$$c \overset{\geq 1, \leq 2}{\rightsquigarrow} d$$

$$d \overset{\geq 1, \leq 1}{\rightsquigarrow} e$$

- Token observation (V)



Example: Transitivity

- Type-level relationships (X)

$$c \overset{\geq 1, \leq 2}{\rightsquigarrow} d$$

$$d \overset{\geq 1, \leq 1}{\rightsquigarrow} e$$

$$\varepsilon_{avg}(d,e) \times P(d_3 \mid c_2, e_4) \times 1$$

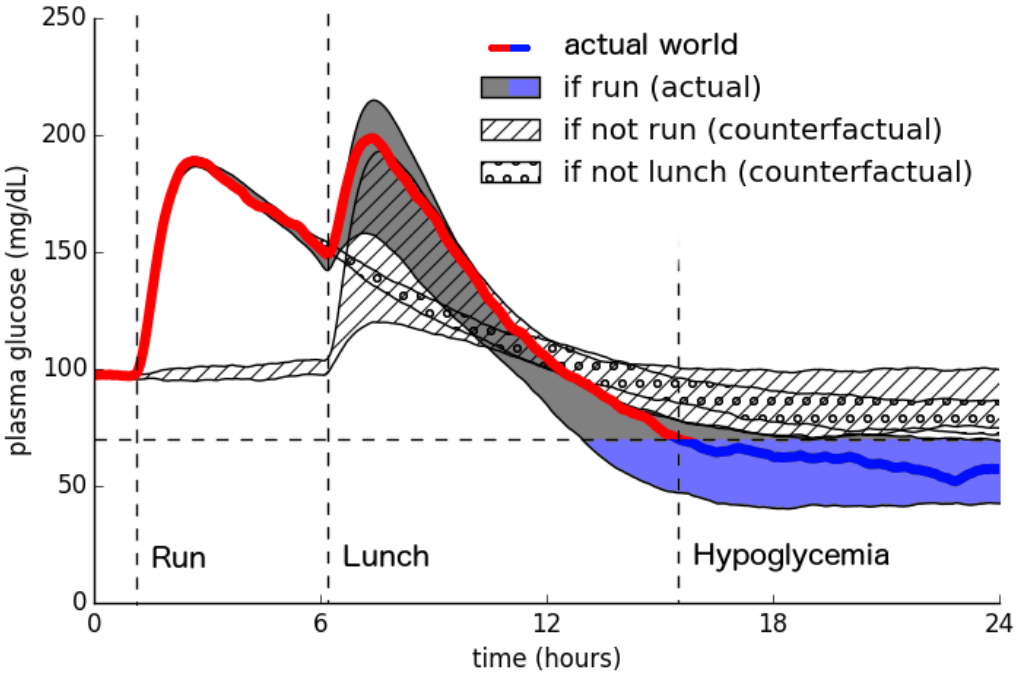
- Token observation (V)



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[glu | R] - E[glu | ¬R]



Wiener

Causality amounts to increase in predictability

Wiener, N. (1956). The theory of prediction. *Modern Mathematics for Engineers. New York: McGraw-Hill*, 165-190.

Granger causality

Basic idea:

Information contained in cause that helps predict effect (which happens after) – that is not contained in other variables

Granger causality - definition

$X_1(t)$ value of X_1 at time t

$X_1^*(t)$ values of X_1 up to time t

$W^*(t)$ all knowledge up to time t

X_2 Granger-causes X_1 if

$$P(X_1(t+1) \mid W^*(t)) \neq P(X_1(t+1) \mid W^*(t) - X_2^*(t))$$

Feedback

X_2 Granger-causes X_1 if

$$P(X_1(t+1) \mid W^*(t)) \neq P(X_1(t+1) \mid W^*(t) - X_2^*(t))$$

Can have X_1 causes X_2 and X_2 causes X_1

Theory versus practice

- Ideal: use all variables over infinitely long timescale
- Reality: limited data, limited computational power necessitate simplifications

bivarate vs. multivariate

- Bivariate: W = X_1, X_2
- Multivariate: W = set of variables

Testing for Granger-causality

X_2 Granger-causes X_1 if

$$P(X_1(t+1) \mid W^*(t)) \neq P(X_1(t+1) \mid W^*(t) - X_2^*(t))$$

Basic approach: regression of X_1 on W. If X_2 terms have nonzero coefficients, then they provide info on X_1.

Vector autoregression (VAR)

$$X(t) = A(1)X(t-1) + A(2)X(t-2)...A(m)X(t-m) + \varepsilon(t)$$

m = model order (how many past values are included)

bivariate VAR

$$X_1(t) = \sum_{j=1}^{m} A_{11}(j) X_1(t-j) + \sum_{j=1}^{m} A_{12}(j) X_2(t-j) + \varepsilon_1(t)$$

$$X_1(t) = \sum_{j=1}^{m} A_{11}(j) X_1(t-j) + \sum_{j=1}^{m} A_{12}(j) X_2(t-j) + \varepsilon_1(t)$$

Testing if X_2 causes X_1

-null hypothesis is that $A_{12}=0$ (no causality)

-test if terms significantly differ from zero

-test if epsilon is reduced with inclusion of X_2

Common cause of two effects

X causes both Y and Z

$$Y(t) = \sum_{j=1}^{m} A_{11}(j)Y(t-j) + \sum_{j=1}^{m} A_{12}(j)X(t-j) + \varepsilon_1(t)$$

$$Y(t) = \sum_{j=1}^{m} A_{11}(j)Y(t-j) + \sum_{j=1}^{m} A_{12}(j)Z(t-j) + \varepsilon_1(t)$$

Causal Chain

X causes Y, Y causes Z

$$Z(t) = \sum_{j=1}^{m} A_{11}(j)Z(t-j) + \sum_{j=1}^{m} A_{12}(j)X(t-j) + \varepsilon_1(t)$$

Directionality

Probability raising is symmetric. Is granger causality?

Multivariate VAR

$$X_V(t) = \sum_{j=1}^{m} A(j) \times X_V(t-j) + \varepsilon_V(t)$$

X_V=vector of size V (values of the V variables)
A=VxV matrix

Complexity

For multivariate:

X is vector of size V (V=number of variables)

A is VxV matrix

m lags means sum over m VxV matrices

For bivariate:

2x2 instead of VxV

Chickens and eggs: Part 1 - chickens

Part 1: Did the Chicken Come First?

The following equation was estimated by OLS:

$$Eggs_t = \mu = \sum_{i=1}^{L} \alpha_i \, Eggs_{t-i} = \sum_{i=1}^{L} \beta_i \, Chickens_{t-i} + \epsilon_t;$$

$H_o : \beta_1 = \ldots \beta_L = 0$ (chickens do not Granger cause eggs).

L = no. of lags	F-statistic	P-value	R^2 of the regression
1	.04	.85	.96
2	1.71	.19	.97
3	1.10	.36	.97
4	.79	.54	.97

Thurman WN, Fisher ME (1988) Chickens, Eggs, and Causality, or Which Came First. American Journal of Agricultural Economics 70(2):237-238.

Chickens and eggs: Part 2 - eggs

Part 2: Did the Egg Come First?

The following equation was estimated by OLS:

$$Chickens_t = \mu + \sum_{i=1}^{L} \alpha_i \, Chickens_{t-i} + \sum_{i=1}^{L} \beta_i \, Eggs_{t-i};$$

$H_o : \beta_1 = \ldots = \beta_L = 0$ (eggs do not Granger cause chichens).

L = no. of lags	F-statistic	P-value	R^2 of the regression
1	1.23	.27	.73
2	10.36	.0002	.81
3	5.85	.0019	.81
4	4.71	.0032	.82

Granger causality vs. causality

Why is it insufficient?

Common causes

Granger causality vs. causality

Why is it unnecessary?

Can have causality without Granger causality

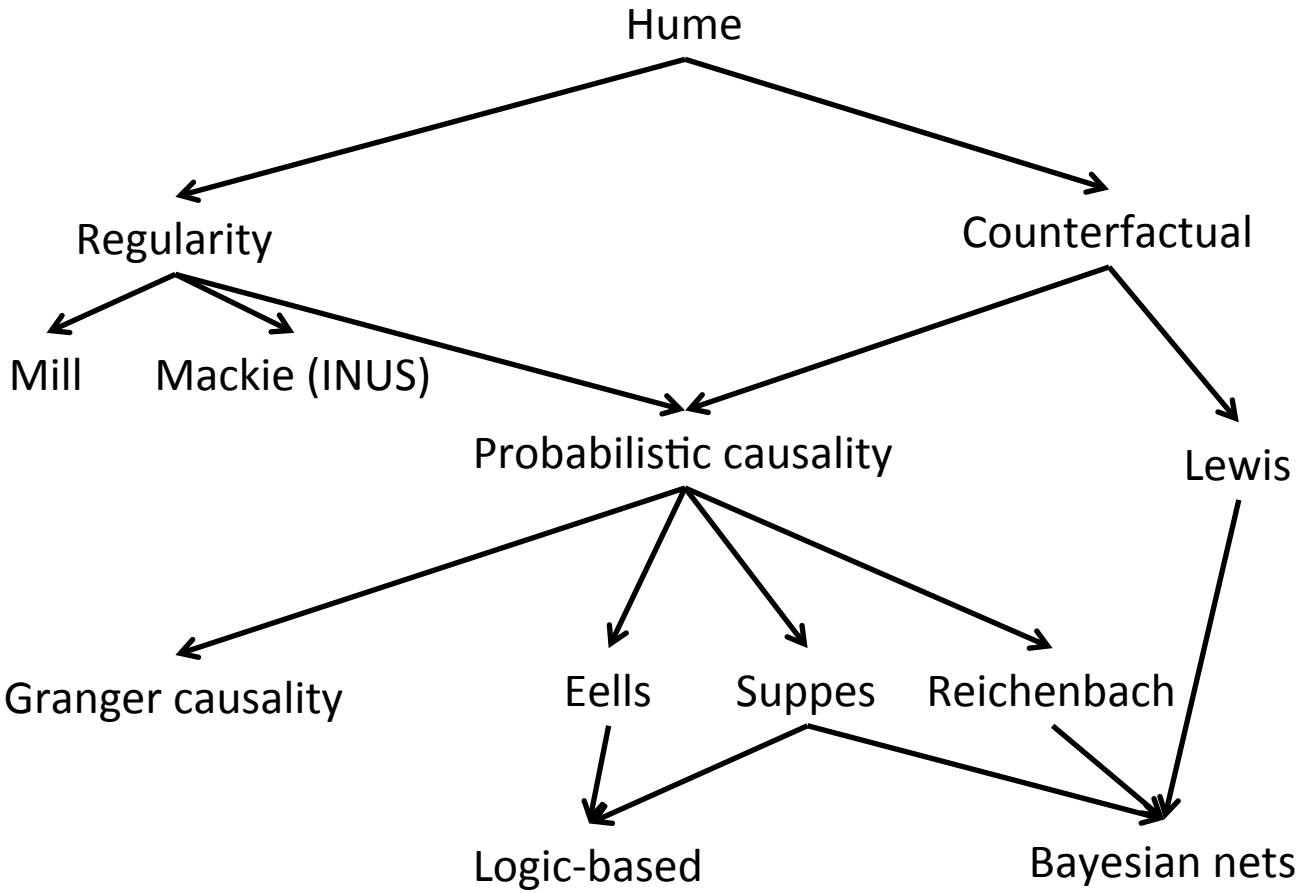
- -Timescale shorter than measured/tested
- -Unfaithful distribution (canceling, control)

Software

- MSBVAR granger.test: bivariate (R package)
- MVGC Multivariate Granger Causality Toolbox (Matlab toolbox): bivariate and multivariate
  - http://www.sussex.ac.uk/sackler/mvgc/

	BN	DBN	Granger	Temporal logic
Results	Graph	Graph	Relationships	Relationships
Time	No	Set of lags	Single lag	Window
Data	C/D/M	C/D/M	C	D/M
Cycles	No	Yes	Yes	Yes
Latent vars	Yes	Yes	No	No
Prediction	Yes	Yes	No*	No
Token cause	Counterfactual-based	No	No	Probabilistic

*For bivariate



Problems

Nonstationarity

Preemption

Overdetermination

Determinism

Causal chains

Recap

- Concepts of causality
  - Regularities, Counterfactuals, Probabilistic
- Two levels
  - Type and token
- Three methods
  - Graphical models (BN, DBN), logic-based, Granger

Discussion example #1

We have a set of data on bike sharing: for each station, when each bike is rented; weather; miles of nearby bike lanes

1) How can we find what causes bike rentals?
2) How can we predict how many bikes will be rented at a particular station on a particular day?
3) What are limitations of this data? Can other datasets help?

https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset

Discussion example #2

You are a researcher at FooBar Corp, which produces widgets.

1) They have collected many years of detailed data on their employees and manufacturing process and want to use this data to improve efficiency (defined as the number of widgets that can be produced each hour).

What approach would you use to analyze their data, why would you choose that, and what information would FooBar Corp. need to best use these results?

2) If FooBar Corp were willing to collect new data, what information would you seek?

Review questions

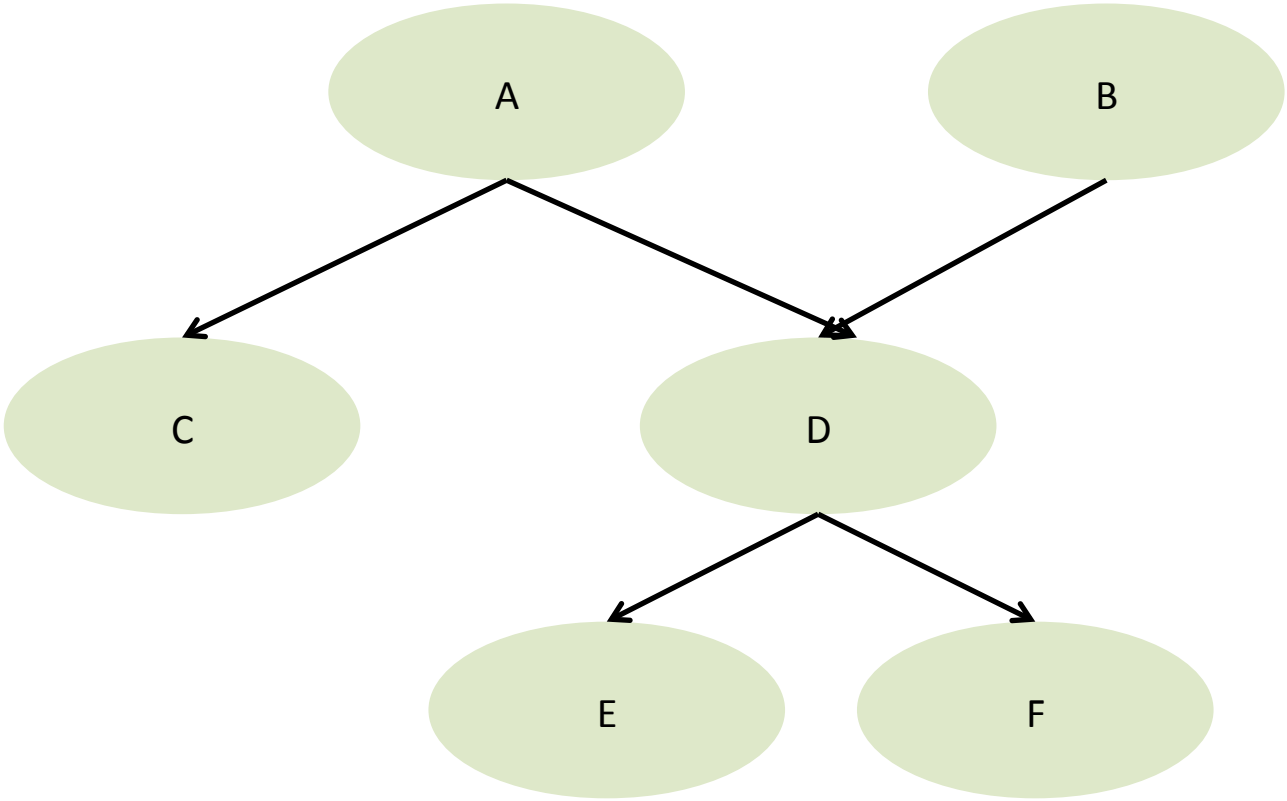
From October 12th New York Times:

Want Your Union to Last? Marry in New Jersey

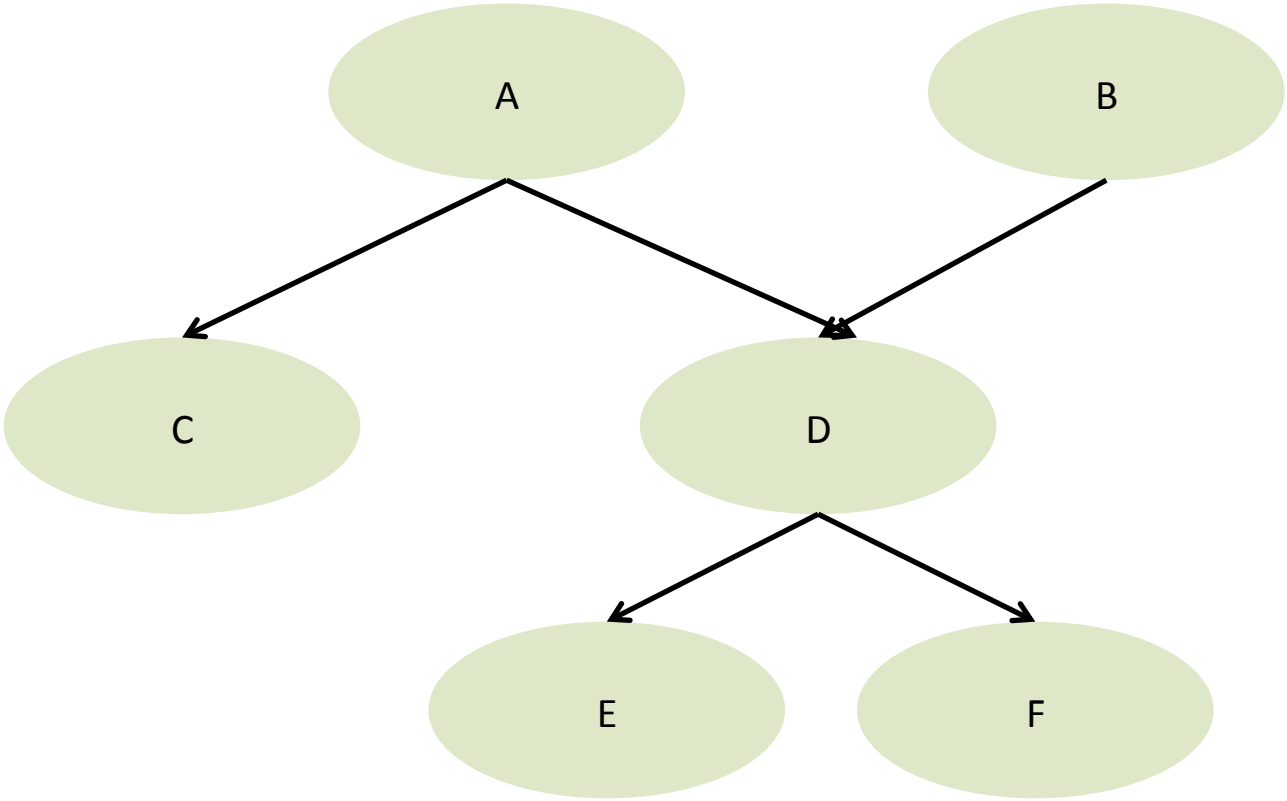
NJ had lower percent of divorced residents over 18 (9%)

http://www.nytimes.com/2012/10/14/fashion/weddings/new-jersey-has-lowest-divorce-rate.html

P(E|do(D))?= P(E|D)



P(E|D,do(A))?=P(E|D)



Ignorability

Bob likes multiple choice exams.

$P(H|M)=0.7$

What's the probability an exam is multiple choice if Bob's happy?

$P(H)=0.2$

$P(M)=0.1$

Answer: 0.35

Does an apple a day keep the doctor away?

Data: days apples eaten, # of doctor visits

Why can’t we go directly from type-level knowledge to token-level cases?

Jane’s study group all developed colds. Jane has been taking a vitamin D supplement and didn’t get a cold. Is the supplement responsible?

Types of problems

- Preemption
- Overdetermination
- Nonstationarity
- Indeterminism
- Hidden variables
- Unfaithful distribution
- Transitivity

Problem cases

- Regularity?
- Counterfactual?
- Probabilistic?

What is the key observation that helps is get causes from probabilities?

Inference in Bayes nets

Approximate vs exact?

Learning graphical models (structure)

Two main approaches?

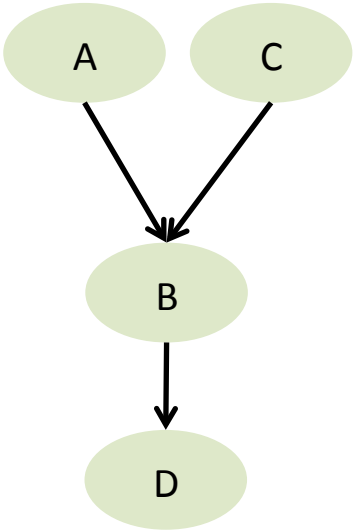
What set of relationships could not be represented with a BN?

What Bayes net(s) is consistent with these independencies?

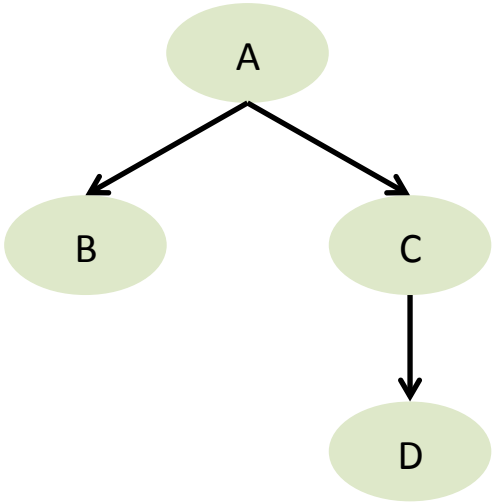
$$A \perp C$$
$$C \perp D \mid B$$
$$A \perp D \mid B$$
$$A \not\perp C \mid D$$

$$A \not\perp B$$
$$C \not\perp B$$
$$D \not\perp B$$

Answer:



What conditional independencies does this graph imply?



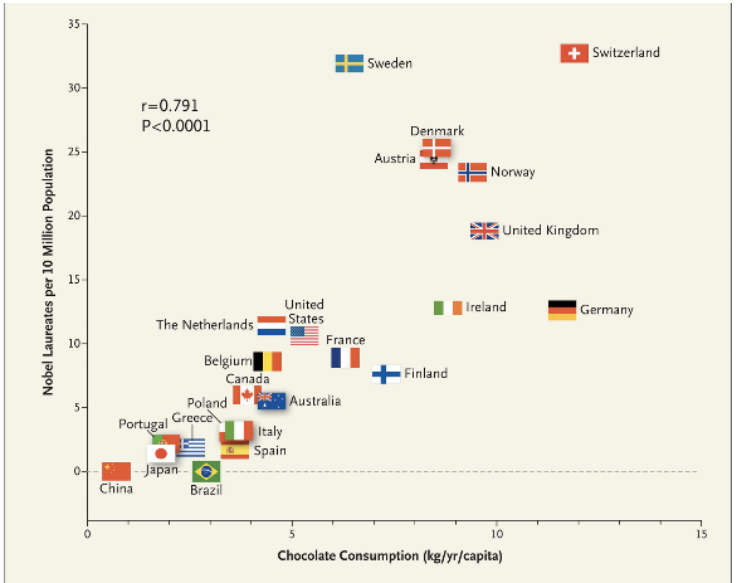
Countries that eat more chocolate win more Nobel prizes. The correlation is statistically significant and chocolate consumption seems to rise at the same rate as Nobel prize winners.

What relationships can explain this?

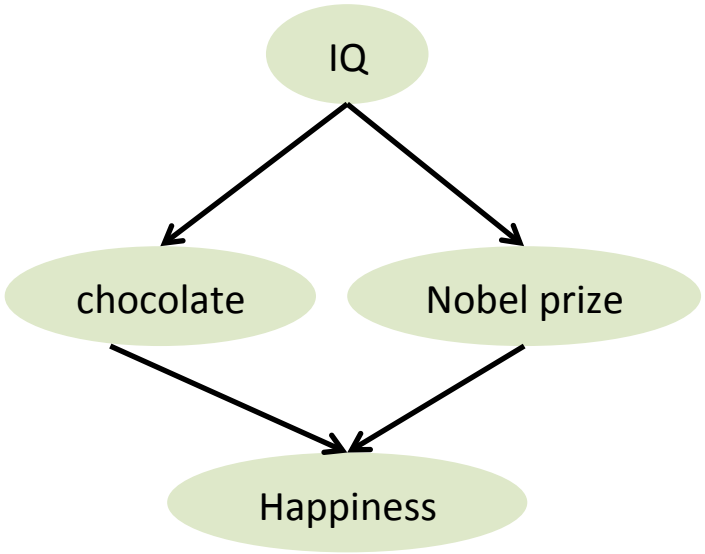
Should we

a) move somewhere with high chocolate consumption?

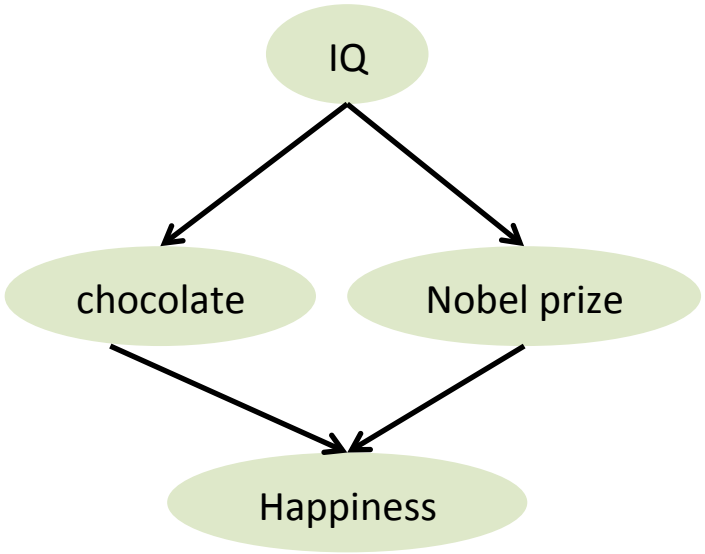
b) stay here but eat more chocolate?



Messerli FH (2012) Chocolate Consumption, Cognitive Function, and Nobel Laureates. N Engl J Med.



If I intervene and give someone chocolate, what is the probability they're happy?



Now say…

H=C^N

C=I

N=I

If someone is happy, will they still be happy if they didn't have chocolate?

For next week: midterm

Two weeks: project proposal