Distinguishing Cause and Effect

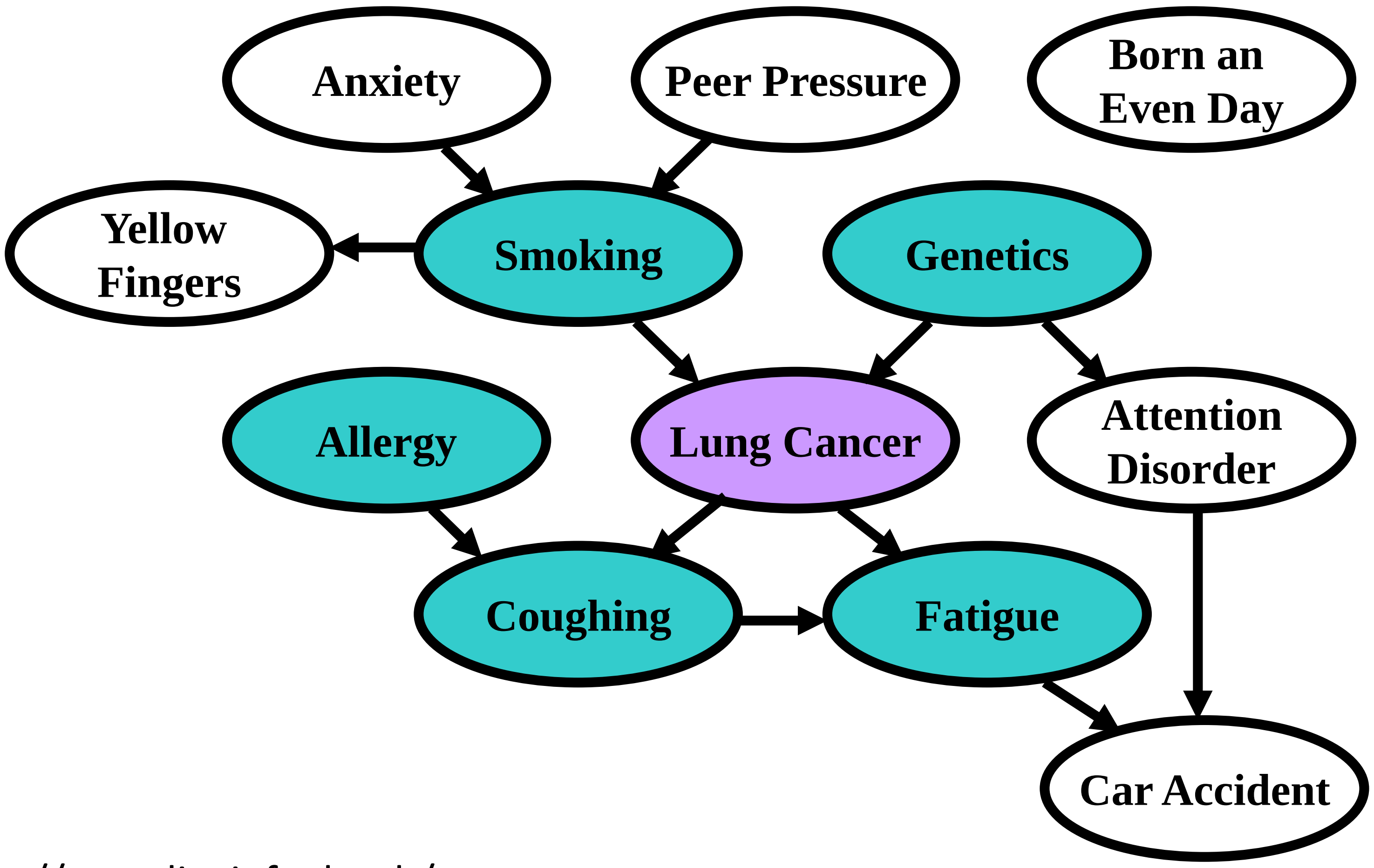
Balram Meena Lohit Jain

Indian Institute of Technology Kanpur

Motivation

- Pervasive in Science, Medicine, Economy and many aspects of everyday life.
- What affects Health, Economy, Climate Changes?
- Gold Standard: Randomized Controlled Experiments
- Experiments Costly, Unethical, Unfeasible!
- Non Observational Routine Data easily available

Causal Graph Example



http://causality.inf.ethz.ch/cause-effect.php?page=data

Causality Challenge #3: Cause Effect Pairs

- Part of IJCNN 2013 contests
- Results discussed in NIPS 2013
- Proceedings: Journal of Machine Learning Research, Workshop and Conference Proceedings (JMLR)

Causality Challenge #3: Cause Effect Pairs

- Challenge: Rank pairs of variables {A, B} to prioritize experimental verifications of the conjecture that A causes B.
- Determine from the joint observation of samples of two variables A and B that A -> B.
- But, “Correlation does not mean Causation”!
- Could be Consequences of a common cause.

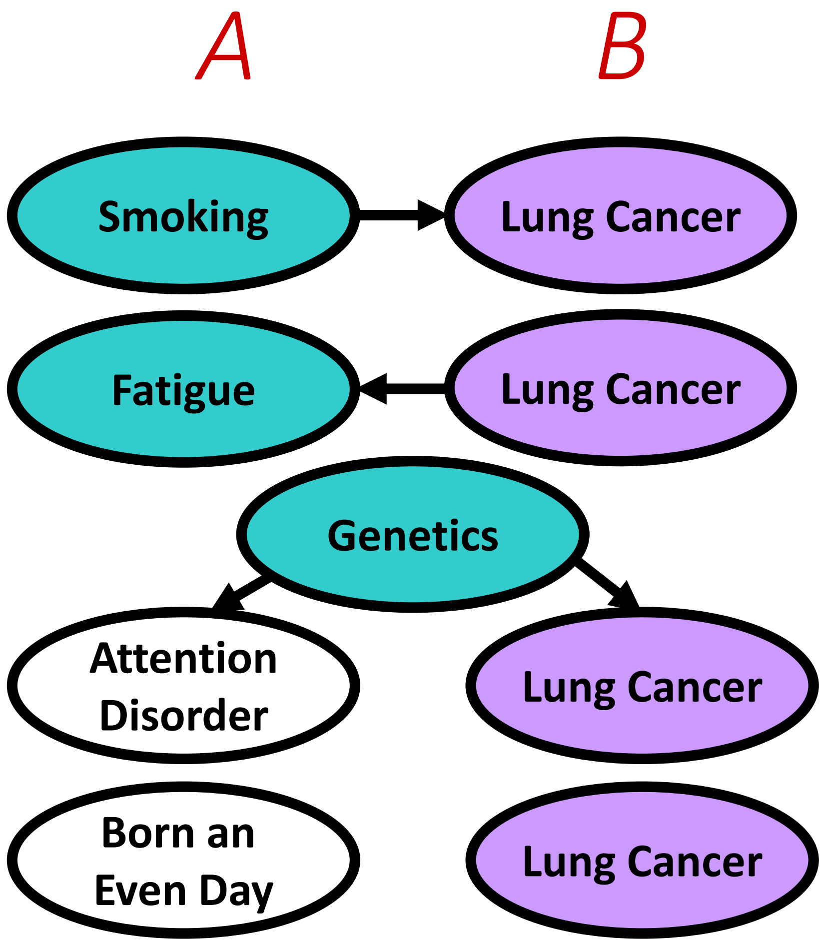
Setup

- No feedback loops.
- No Explicit time information
- Variables are aggregate statistic, eg: Temp, life expectancy.
- Pairs independent of each other

Datasets

- Pair of real variables intermixed with
  - controls (dependent but not causally related) and
  - semi-artificial cause-effect pairs (real variables mixed in various ways to produce a given outcome)
- 4050 training pairs
- 4050 validation pairs
- 4050 test pairs

Cause Effect Pair problem



A -> B

A <- B

A – B

$A \perp B$

http://causality.inf.ethz.ch/cause-effect.php?page=data

Evaluation Scheme

- For any pair, score between -Inf and +Inf,
- Large positive values : A is a cause of B with certainty
- Large negative values : B is a cause of A with certainty
- Near zero : Neither A causes B nor B causes A
- Scores as ranking criterion
- Evaluate entries with two *Area under the ROC Curve* (AUC) scores

Area Under the ROC curve

- The results of classification, obtained by thresholding the prediction score, may be represented in a confusion matrix, where tp (true positive), fn (false negative), tn (true negative) and fp (false positive) represent the number of examples falling into each possible outcome:
- We define the sensitivity (also called true positive rate or hit rate) and the specificity (true negative rate) as:
  - Sensitivity = tp/pos
  - Specificity = tn/neg

  where pos=tp+fn is the total number of positive examples and neg=tn+fp the total number of negative examples.
- The area under the curve obtained by plotting sensitivity against specificity by varying a threshold on the prediction values to determine the classification result.
- The AUC is calculated using the trapezoid method.

Causality in two variables : Intuitively

- Intuitively : Factorization of the joint distribution

  P(cause; effect) into P(cause)P(effect | cause)

  typically yields models of lower total complexity than

  P(cause; effect) into P(effect)P(cause | effect)

- Definition of Notion of Intuition not obvious!

Previous Models

- The methods define classes of conditionals ***C*** and marginal distributions ***M***, and prefer
- X -> Y whenever P(X) ∈ M and P(Y | X) ∈ C

  but P(Y) ∉ M or P(X | Y) ∉ C.
- Notion of model complexity: all probability distributions inside the class are simple, and those outside the class are complex.
- This a priori restriction poses serious practical limitations

Causality in two variables

- Deterministic $f(X,E) = F(X)$
- Non-deterministic
  I. AN(additive noise) $f(X,E) = F(X) + E$
  II. PNL (Post-Non-Linear model) $f(X,E) = G(F(X) + E)$
  III. LINGAM (f is linear) $f(X,E) = pX + qE$
  IV. HS (hetro-Schedastic noise) $f(X,E) = F(X) + E.G(X)$
- Idea is to fit restriction model in both direction (X -> Y and Y -> X)
- Direction to be one that yields the best fit.

Probabilistic Latent Variable : Additional Assumptions

A. Determinism (no other causes of Y): a function f exists such that $Y = f(X,E)$
B. X and E are independent.
C. The distribution of the cause is "independent" from the causal mechanism (f)
D. The noise has a standard-normal distribution: $E \sim N(0,1)$

Other Models

- Based on (A) and (B) with some additional restrictions on f (Slide 13).
- For these special cases, it has been shown that a model of the same (restricted) form in the reverse direction Y -> X that induces the same joint distribution on (X, Y) does not exist in general.
- But, a limited model class may lead to wrong conclusions about the causal direction.

Probabilistic Latent Variable Model

- In general, one can always construct a random variable E' ~ N(0,1) and a f' : R^2 -> R such that

$$X = f'(Y, E')$$

- In combination with (C) and (D) : an asymmetry!
- Infer the causal direction

Basic Idea

- Define non-parametric priors on the f and input distributions favoring lower complexity.
- Inferring using standard Bayesian model selection
- Preference to model with largest marginal likelihood
- Bayesian Approach: Noise as Latent Variable summarizing influence of all other unobserved causes.

Bayesian Model Selection

- Prefer model with highest evidence:

$$\rho(D|M) = \int \rho(D|\theta, M)\rho(\theta|M)d\theta,$$

D=Data, M=Model, θ=Parameters

Trade-off between likelihood (goodness of fit) and priors (model complexity).

- Causal Discovery: Compare evidence X->Y and Y->X

References

- Mooij, Joris M., et al. "Probabilistic latent variable models for distinguishing between cause and effect." NIPS. 2010.
- Daniusis, Povilas, et al. "Inferring deterministic causal relations." arXiv preprint arXiv:1203.3475 (2012).
- Hoyer, Patrik O., et al. "Nonlinear causal discovery with additive noise models." NIPS. Vol. 21. 2008.
- Peters, Jonas, Dominik Janzing, and Bernhard Scholkopf. "Causal inference on discrete data using additive noise models." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.12 (2011): 2436-2450.
- Janzing, Dominik, et al. "Information-geometric approach to inferring causal directions." Articial Intelligence 182 (2012): 1-31.

Thank You!

Questions …