Beyond Correlation
Counterfactual Reasoning & Causal Inference

Tanmayee Narendra

IBM Research

Karl Pearson

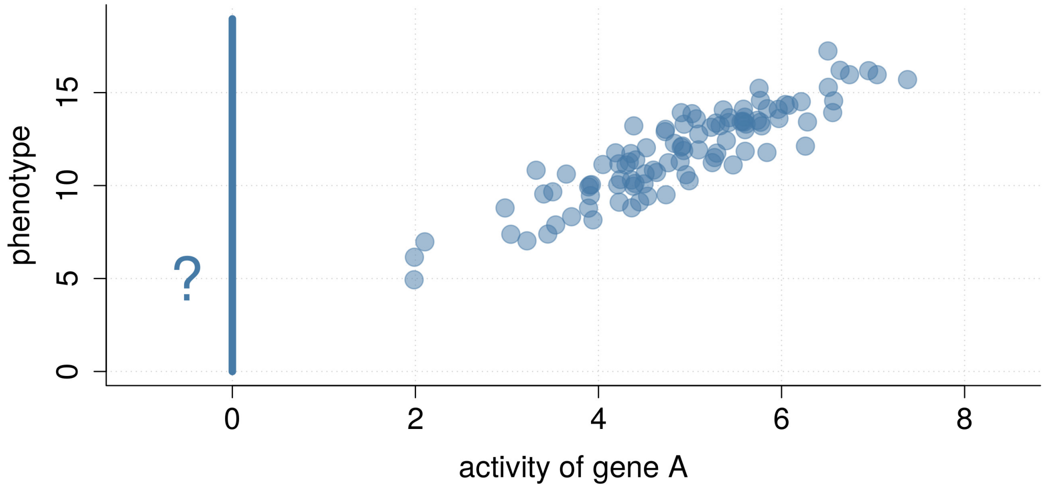
- "Beyond such discarded fundamentals as 'matter' and 'force' lies still another fetish amidst the inscrutable arcana of modern science, namely, the category of cause and effect."
- He categorically denied the need for an independent concept of causal relation beyond correlation.

Porter, Theodore M. "Karl Pearson." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 20 Mar. 2018, www.britannica.com/biography/Karl-Pearson.

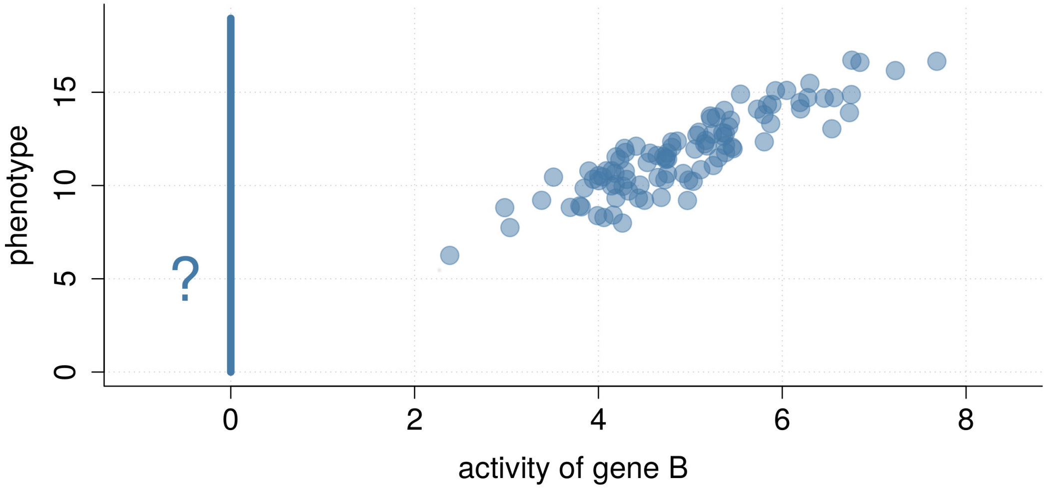
What is Causality? Why do we need it?

An Example

Gene A



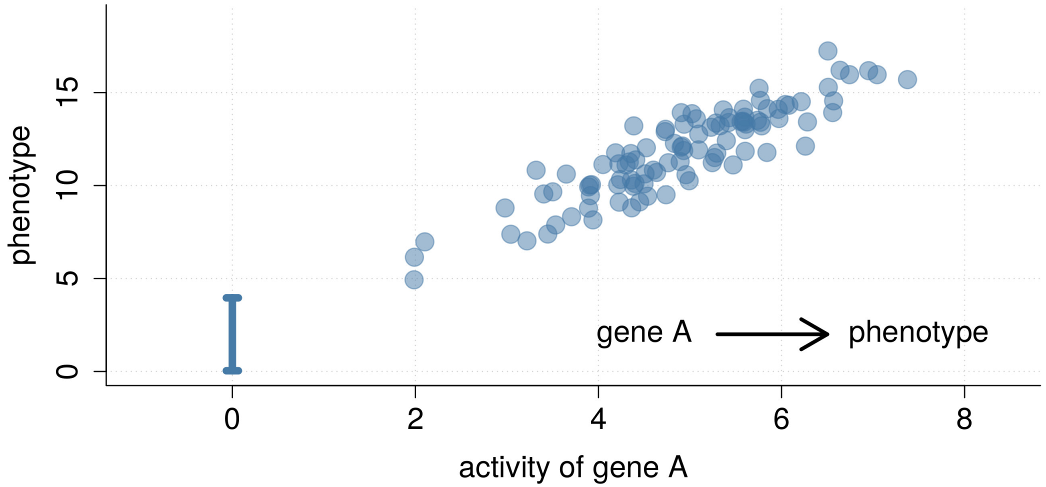
Gene B



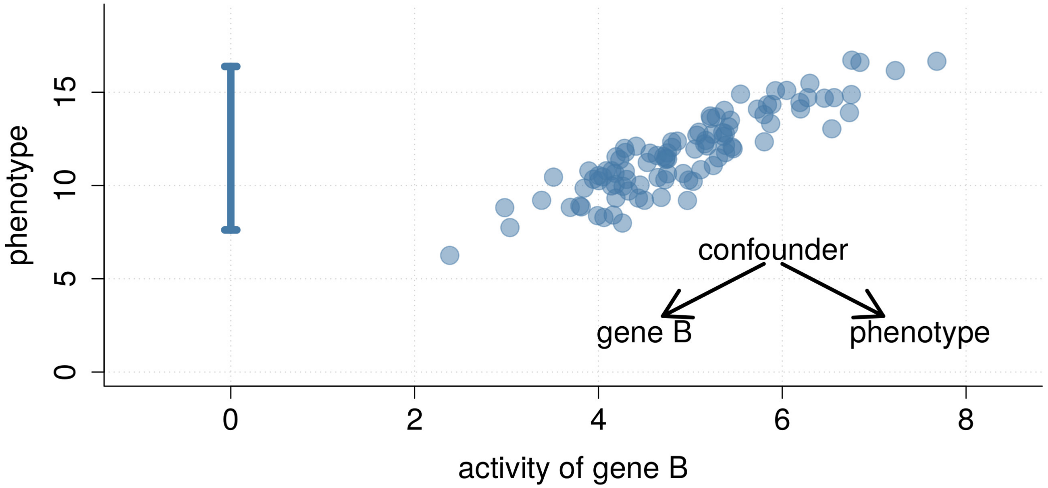
Adapted from Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT Press, 2017.

An Example

Gene A

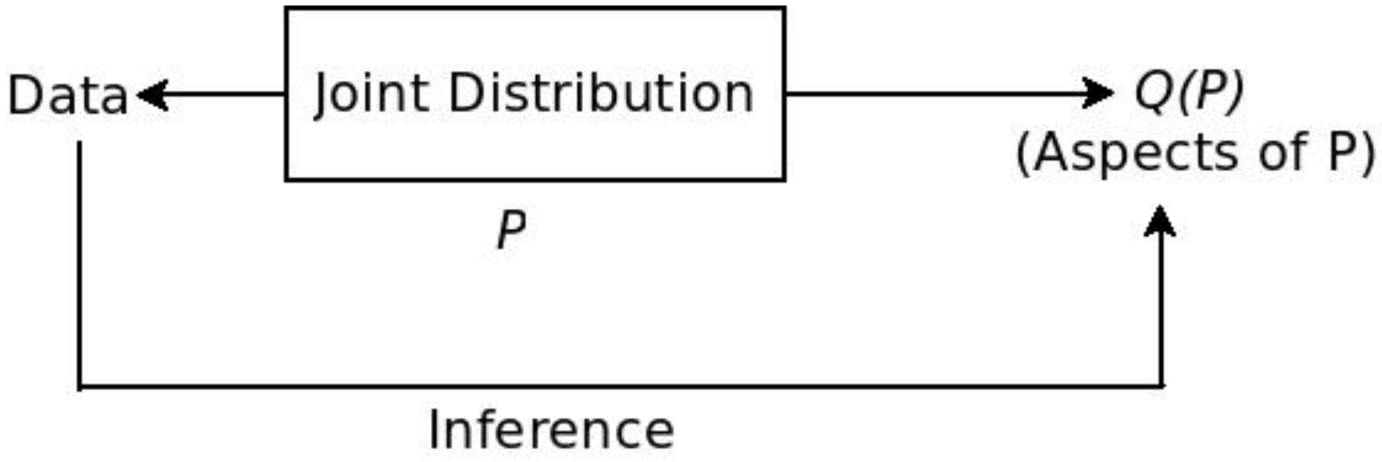


Gene B



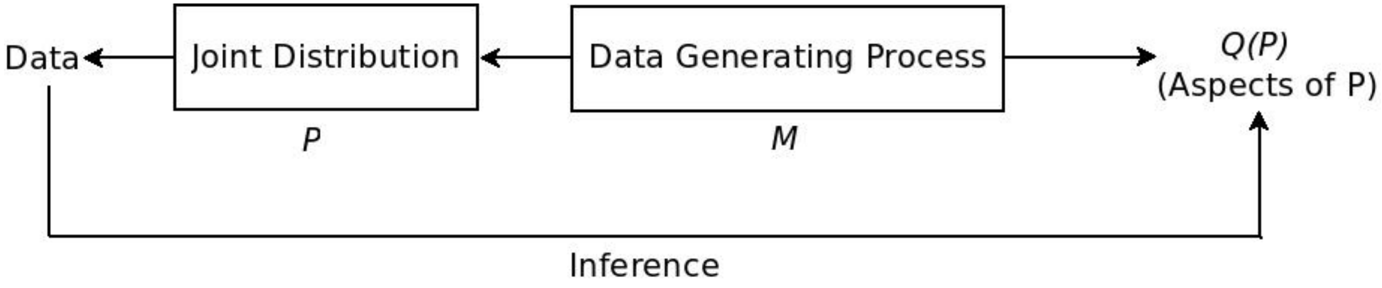
Adapted from Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT Press, 2017.

Statistical Inference Paradigm



Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Causal Paradigm



Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Causal Model

- Any causal model must be able to answer the following types of questions
  1. Observational questions
     What if we *see* A?
  2. Action questions (interventions)
     What if we *do* A?
  3. Counterfactual questions
     What if we *did things differently*?
- Parametric and Non-parametric

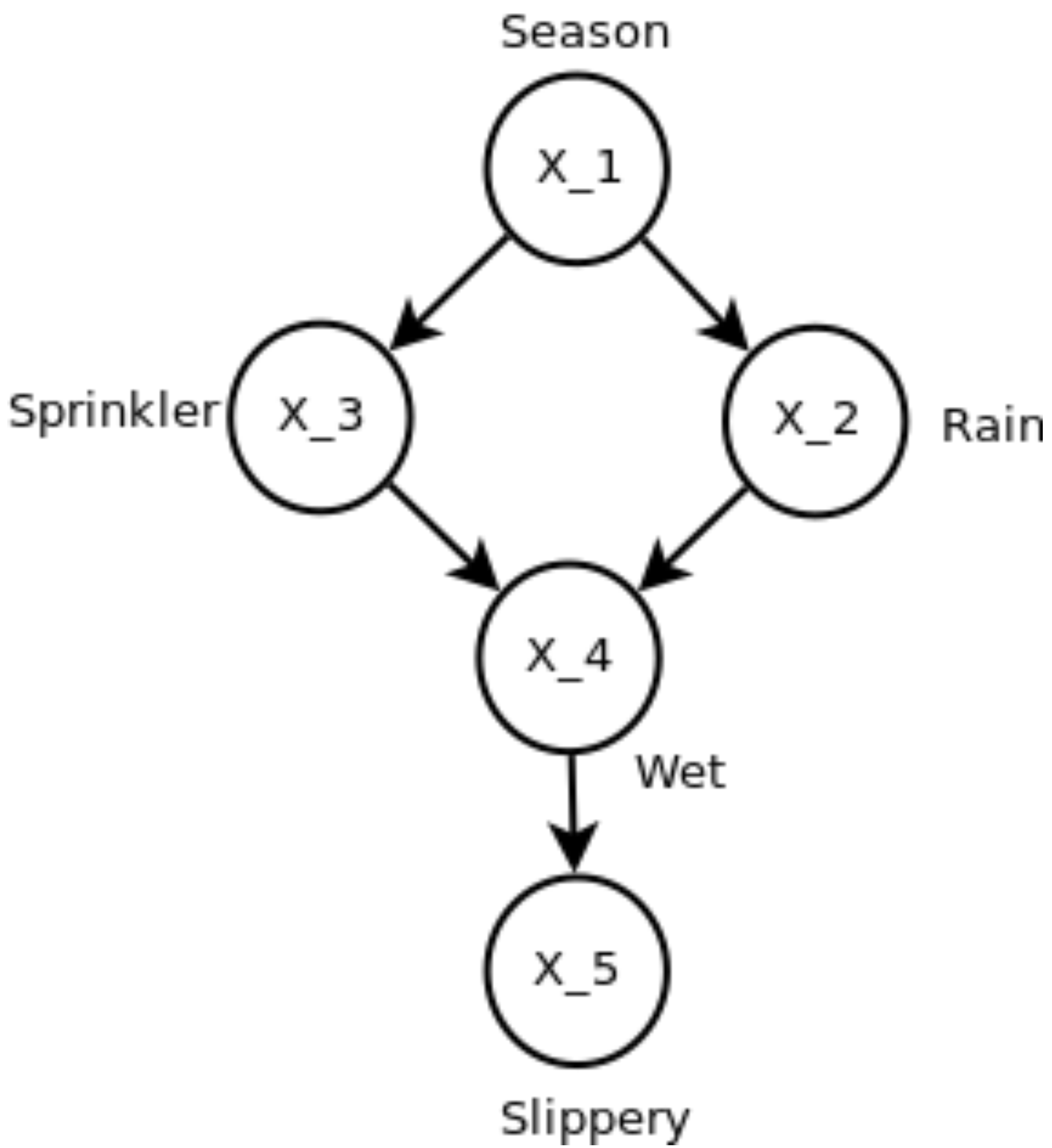
Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Causal Model

- Any Causal Model usually comprises of
  - Causal Graph
  - Distribution
  - Intervention Distributions
  - Counterfactuals

Prediction

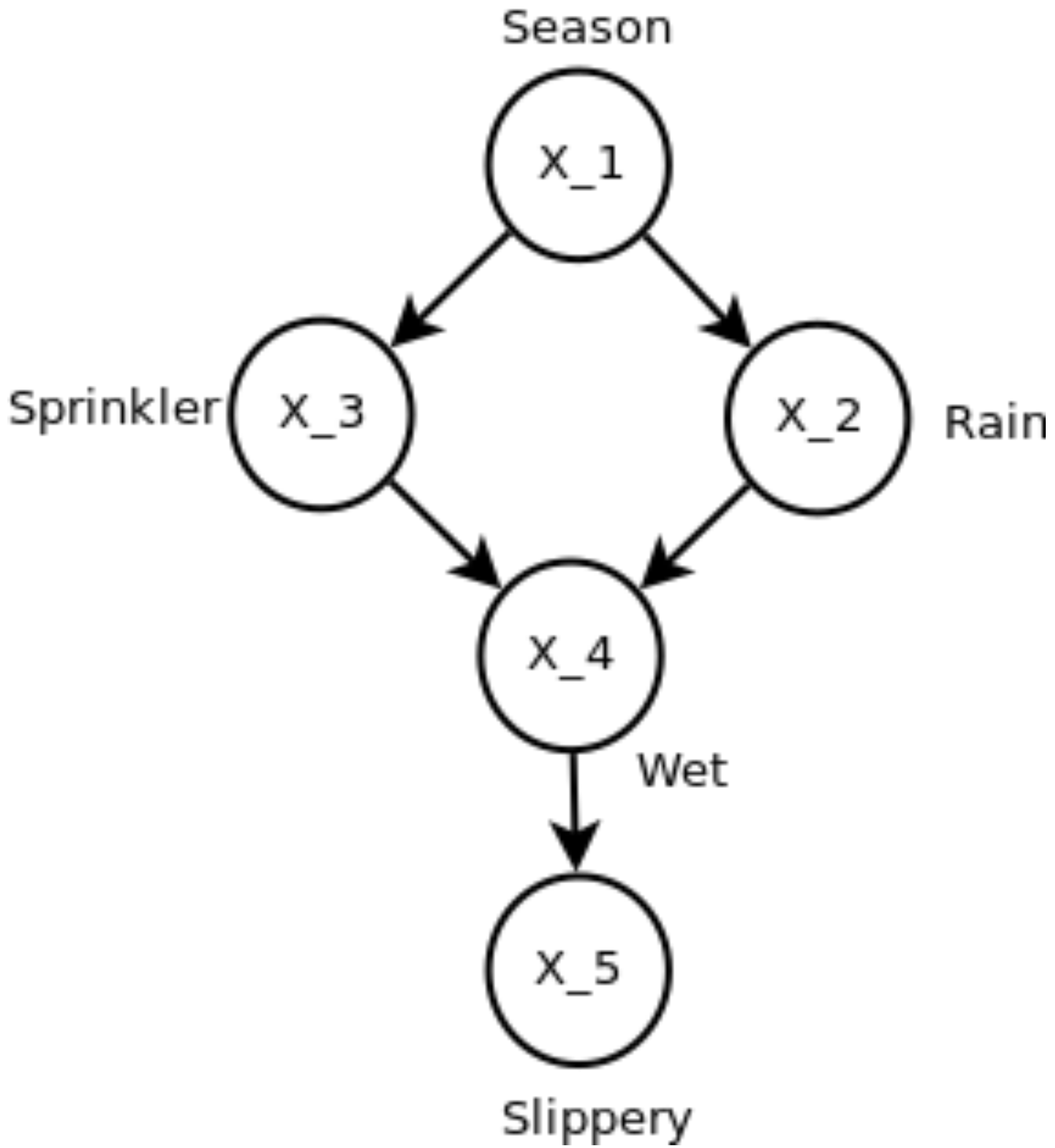
Would the pavement be slippery if we find the sprinkler off?



Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Intervention

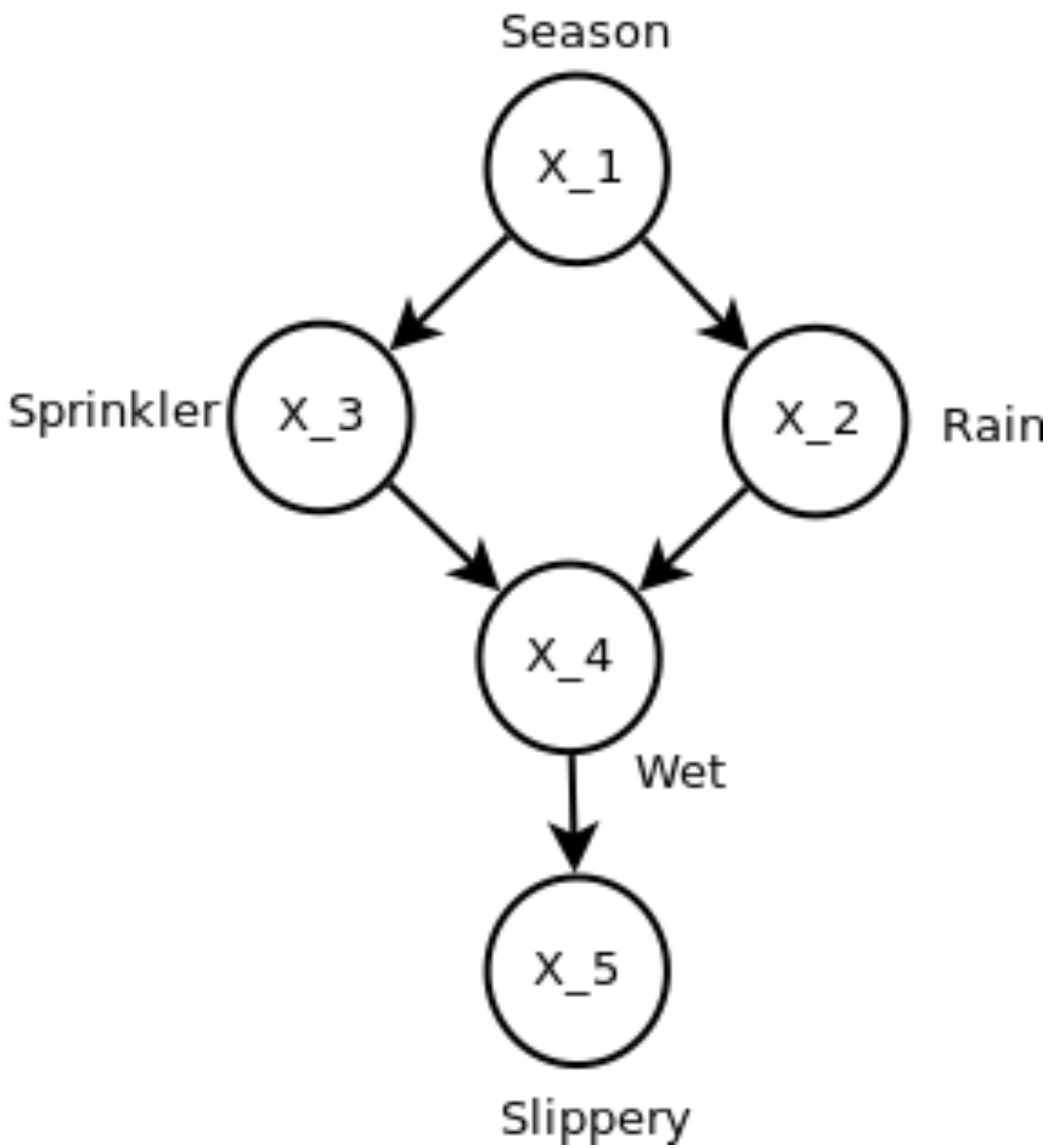
Would the pavement be slippery if we make sure the sprinkler is off?



Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Counterfactual

Would the pavement be slippery had the sprinkler been off, given that the pavement is in fact not slippery and the sprinkler is on?



Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Causal Bayesian Networks

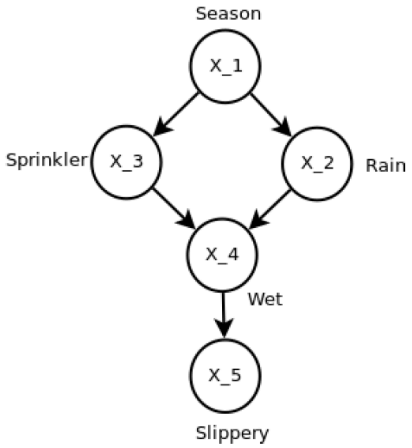


Figure: Causal Bayesian Network

$$Pr(X) = Pr(X_1)\, Pr(X_2 \mid X_1)\, Pr(X_3 \mid X_1)\, Pr(X_4 \mid X_3, X_2)\, Pr(X_5 \mid X_4)$$

Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Intervention

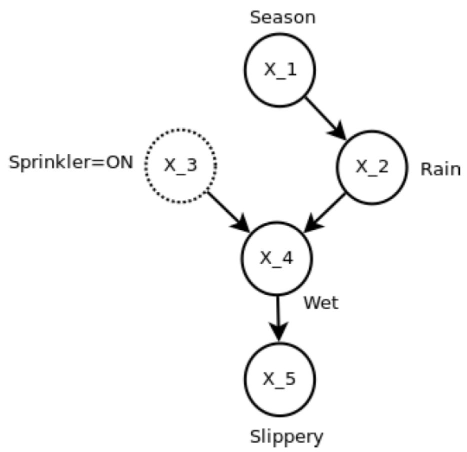
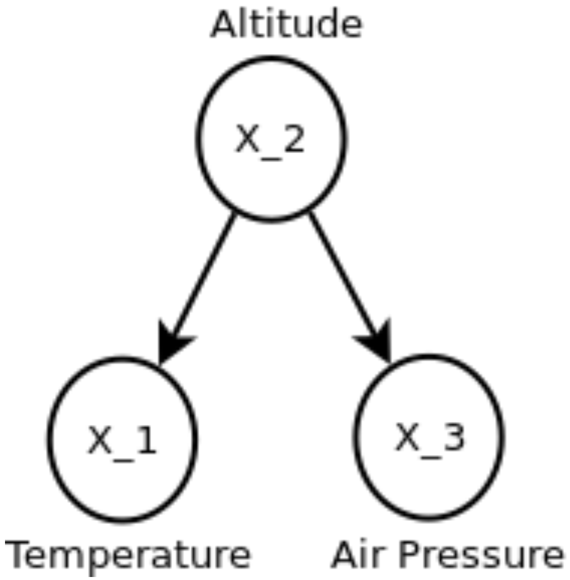


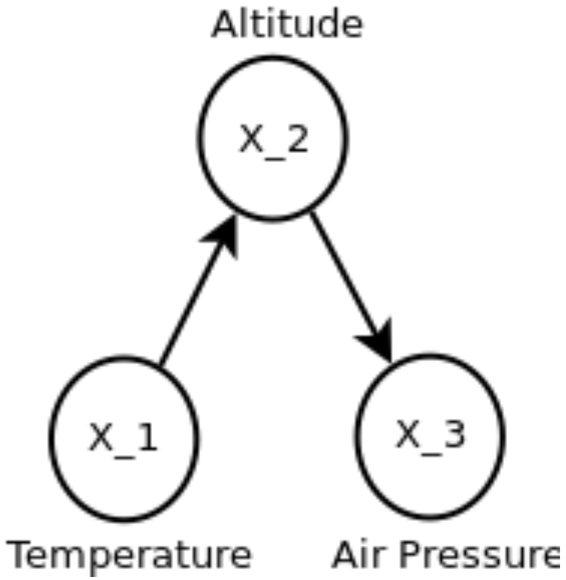
Figure: Causal Bayesian Network

$$Pr(X)_{X_3=ON} = Pr(X_1)\,Pr(X_2 \mid X_1)\,Pr(X_4 \mid X_3 = ON, X_2)\,Pr(X_5 \mid X_4)$$

Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Markov Equivalent Class





Adapted from Pearl, Judea. *Causality*. Cambridge university press, 2009.

Causal Bayesian Networks

Definition

Let $\mathcal{L}(\mathbf{X})$ be a probability distribution on a set $\mathbf{X}$ of variables, and let $\mathcal{L}_\nu(\mathbf{X})$ denote the distribution resulting from the intervention $do(V = \nu)$ that sets a subset V of variables to constants ν. Denote by $\mathcal{L}*$ the set of all interventional distributions $\mathcal{L}_\nu(\mathbf{X})$, $V \subseteq \mathbf{X}$, including $\mathcal{L}(\mathbf{X})$, which represents no intervention (i.e., $X = \phi$).

Adapted from Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT Press, 2017.

Causal Bayesian Networks

Definition

A DAG G is said to be a causal Bayesian network compatible with $\mathcal{L}*$ if and only if the following three conditions hold for every $\mathcal{L}_\nu \in \mathcal{L}*$:

1. $\mathcal{L}_\nu(\mathbf{X})$ is Markov relative to G
2. $\mathcal{L}_\nu(X_i) = 1$ for all $X_i \in \mathbf{X}$ whenever X_i is consistent with $V = \nu$
3. $\mathcal{L}_\nu(X_i \mid Pa_i) = \mathcal{L}(X_i \mid Pa_i)$ for all $X_i \notin \mathbf{X}$ whenever Pa_i is consistent with $V = \nu$ i.e., each $\mathcal{L}(X_i \mid Pa_i)$ remains invariant to interventions not involving X_i.

Adapted from Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT Press, 2017.

Structural Equation Models

Causal Bayesian Networks can also be formulated as Structural Equation Models (SEM).

Definition

A structural equation model is defined as a tuple $\mathcal{S} := (\mathbf{S}, \mathbb{P}^{\mathbf{N}})$, where $\mathbf{S} = (S_1, \ldots, S_p)$ is a collection of p equations

$$S_j : \quad X_j = f_j\left(Pa_j, N_j\right), \qquad j = 1, \ldots, p$$

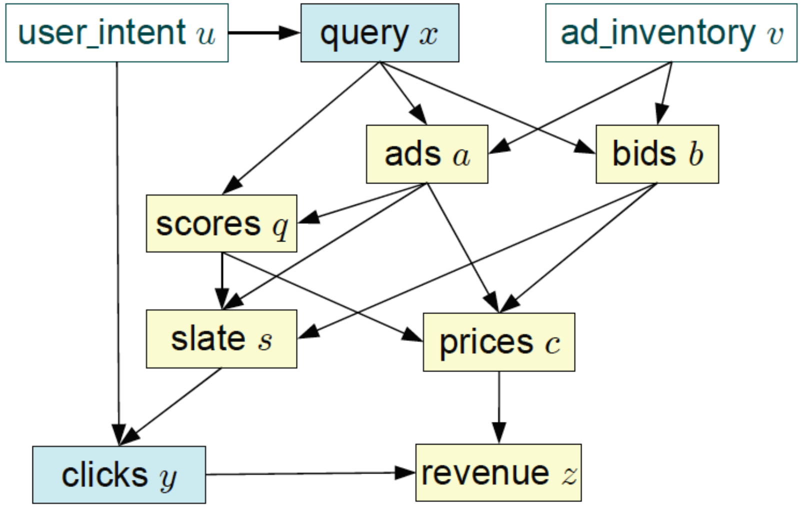
where $\mathbb{P}^{\mathbf{N}} = \mathbb{P}^{N_1,\ldots,N_p}$ is the joint distribution of the noise variables, which are required to be jointly independent.

Adapted from Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT Press, 2017.

Is Causality useful?

Computational Advertising

- Complex system with several ML components, and actors with varied interests
- Traditionally modelled as Contextual Bandits
- Causal Modelling helps in design of the system, by making it principled and cheaper



Bottou, Léon, et al. "Counterfactual reasoning and learning systems: The example of computational advertising." *The Journal of Machine Learning Research* 14.1 (2013): 3207-3260.

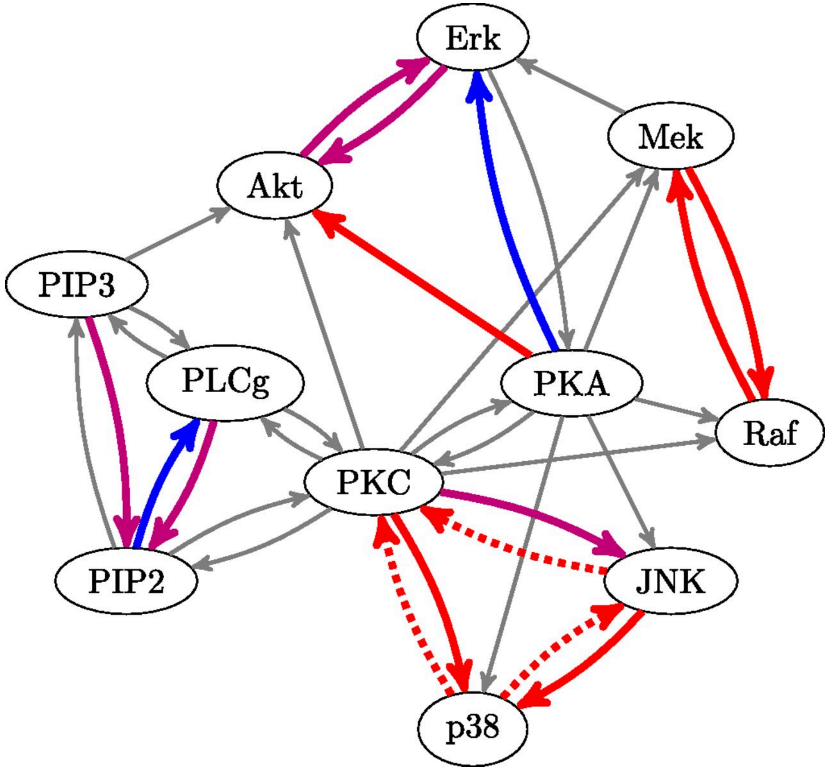
Exoplanet Search

- Removing systemic noise from observations of the Kepler space observatory
- Systemic noise introduced from spacecraft and telescope (pointing jitter)
- New technique called Half-Sibling Regression

Schölkopf, Bernhard, et al. "Removing systematic errors for exoplanet search via latent causes." *International Conference on Machine Learning*. 2015.

Gene Perturbation Experiments

- Improving experimental interventions like gene deletion
- Estimate causal relations between biochemical agents



Peters, Jonas, Peter Bühlmann, and Nicolai Meinshausen. "Causal inference by using invariant prediction: identification and confidence intervals." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78.5 (2016): 947-1012.

Estimating the Effect of a Market Intervention

- Did a particular advertising campaign increase product sales?
- Bayesian Structural Time Series Model
- Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, 2015, Vol. 9, No. 1, 247-274.

Brodersen, Kay H., et al. "Inferring causal impact using Bayesian structural time-series models." *The Annals of Applied Statistics* 9.1 (2015): 247-274.

App Store Analysis

- Estimate which app release is successful and which is not
- Simple application of Causal Impact paper

Martin, William. "Causal impact for app store analysis." *Software Engineering Companion (ICSE-C), IEEE/ACM International Conference on*. IEEE, 2016.

People in Causal Inference

- UCLA - Judea Pearl
- CMU - Peter Spirtes, Clark Glymour, Richard Scheines
- Harvard - Donald Rubin
- ETH-Zurich - Jonas Peters, Peter Buhlmann, Nicolai Meinshausen, Stefen Bauer
- MPI-Tubingen - Dominik Janzing, Bernhard Scholkopf
- Others - Joris Mooij, Patrik Hoyer and many others

Further Reading

- Pearl, Judea. *Causality*. Cambridge university press, 2009.
- Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT Press, 2017.
- Lectures on Causality https://youtu.be/zvrcyqcN9Wo
- And many more

Visit

triptoes1.github.io

Twitter

@tanmayee_n