Causal Inference and Invariance

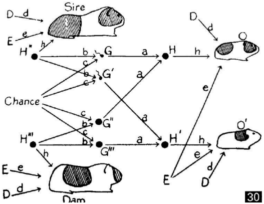
Charles Zheng and Qingyuan Zhao

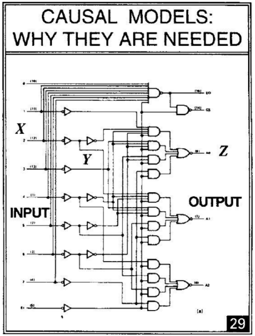
Stanford University

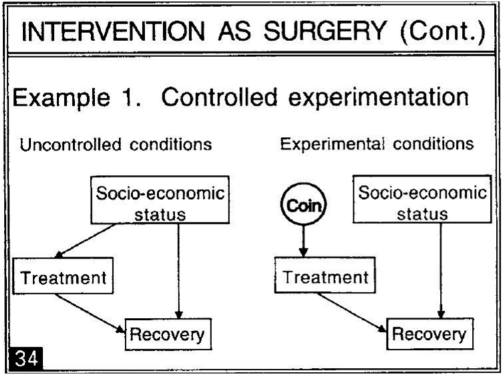
February 12, 2016

(Part 1/2)

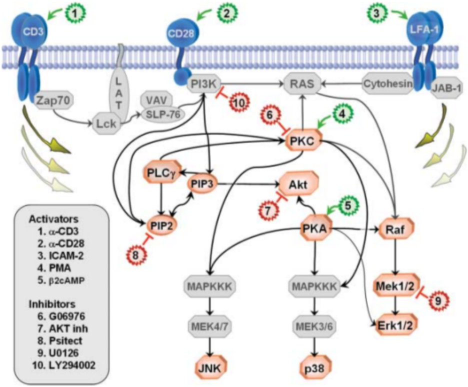
Understanding nature = cause and effect





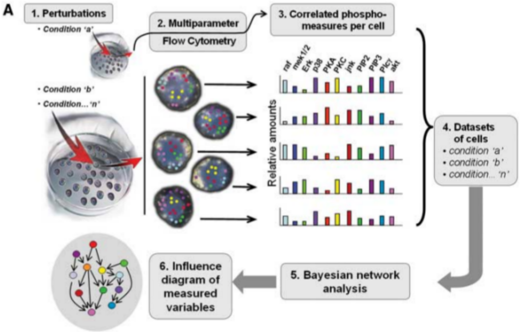


A hot application: systems biology



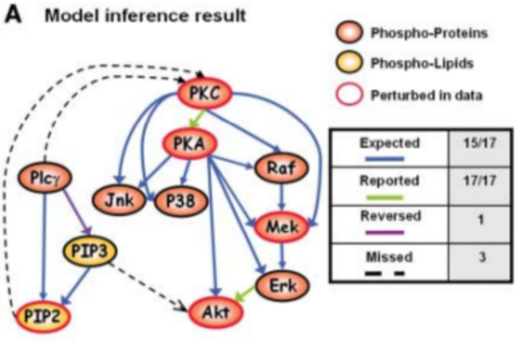
- In biology: causal relationships due to *chemical interactions*.
- Experimenters *intervene* by injecting *activators* and *inhibitors*.

Protein signalling data



- Flow cytometry data from Sachs et al.*Science*, 2005.
- 1 observational data set + 9 interventions (selective activation/inhibition).

Putative causal model



- Causal inference applied to observational + interventional data.
- Recovered most of the known interactions.

Where does statistics fit into this?

Not just statistics: numerous fields are interested in causal inference.

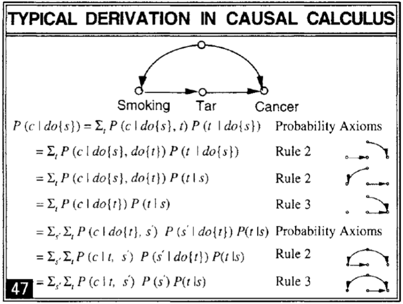
- *Philosophy.* Starting from Aristotle and still being debated today. What is causality? How do we learn about cause and effect?
- *Computer science.* Can we build an artificial intelligence which reasons like humans? Motivation for Judea Pearl's work.
- *Social science.* What influences an individual's life choices? (Career, political participation, etc.) Can we discover *social mechanisms*?

Statistical applications focus on:

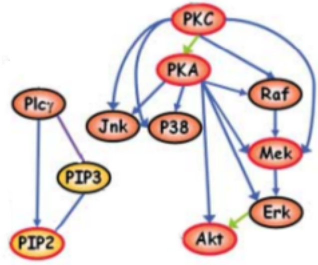
- *Estimating causal effects.* Can we predict a causal effect based on observational or experimental data? E.g. effect of a medical treatment based on clinical trial data? Motviation for potential outcomes approach developed by Rubin, etc.
- *Bayesian networks, structure learning.* Can we model multivariate relationships using a network structure? Networks *can be* given causal interpretation, but causal inference is not the only motivation. Motivation for graphical lasso.

Principles of Causal Inference

- These diverse applications of causal inference share a common set of useful principles.
- The *graphical approach* pioneered by Judea Pearl is the best approach for developing causal intuition.

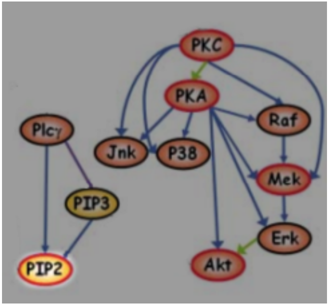


Graphs: nodes and vertices



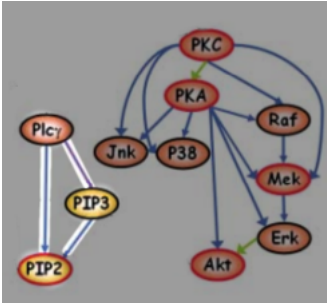
- Each variable in the dataset is given a *node*.
- Arrows indicate which variables *cause* which other variables.
- Undirected or bidirected edges indicate that variables are associated, but neither causes the other.

Causality and experiments



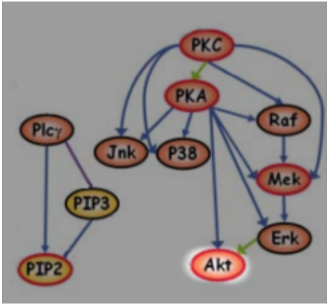
- The *observational distribution* is the joint distribution under the "natural" state of the system.
- One can consider *intervening* on one of the variables in the system. This changes the joint distribution of the system.

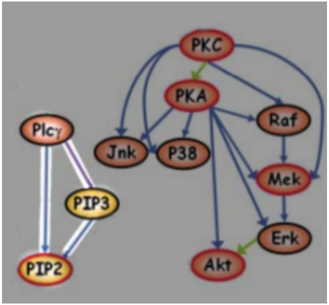
Causality and experiments



- However, not every variable will be affected by the intervention!
- By following the arrows, we determine the set of variables which are affected by the intervention.

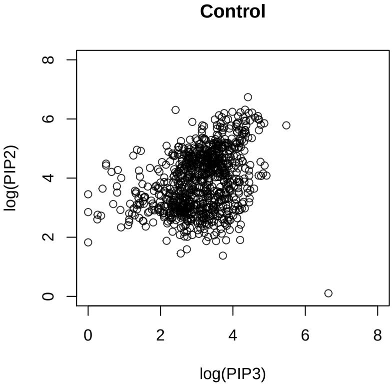
Principle I: Which variables are affected.





- If we *inhibit* Akt, no other variables should be affected.
- If we *inhibit* PIP2, then we may not only change the distribution of PIP2, but also PIP3.

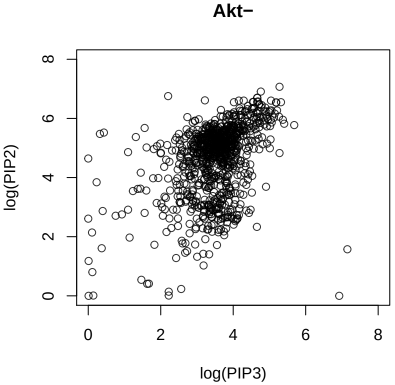
Principle I: Which variables are affected.



Looking at Sachs data.

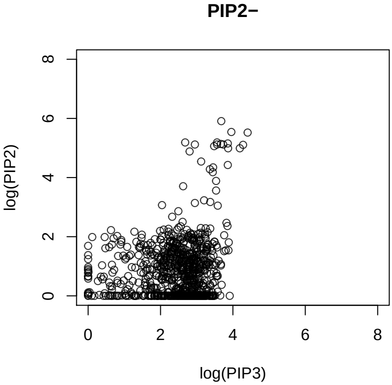
Joint distribution of PIP2 and PIP3 in the “control” case.

Principle I: Which variables are affected.



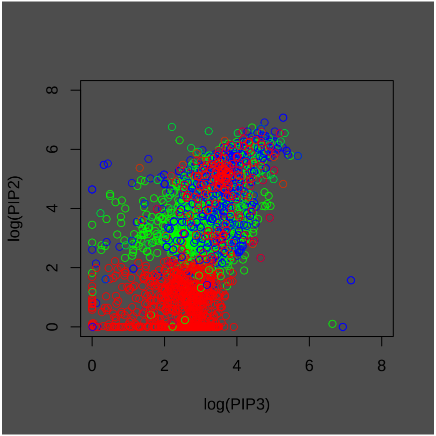
Joint distribution of PIP2 and PIP3 when we intervene on Akt.

Principle I: Which variables are affected.



Joint distribution of PIP2 and PIP3 when we intervene on PIP2.

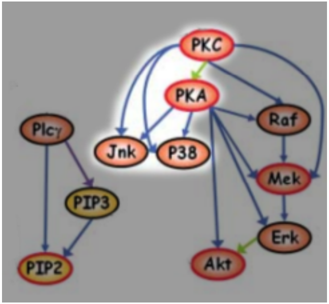
Principle I: Which variables are affected.



Control, **PIP2-**, **Akt-**

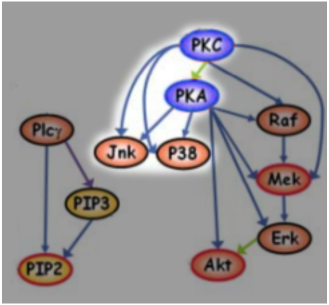
Intervening on PIP2 also affects the distribution of PIP3, while intervening on Akt does not (drastically) change the distribution.

Principle II: Conditional independence.



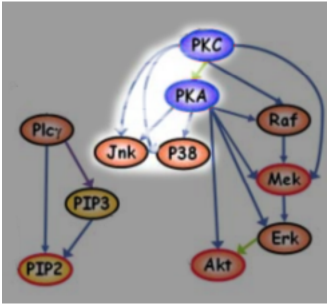
- Surprisingly, the structure of the causal graph implies certain *conditional independence* relationships.
- This allows the potential to infer causal relationships from observational data.

Principle II: Conditional independence.



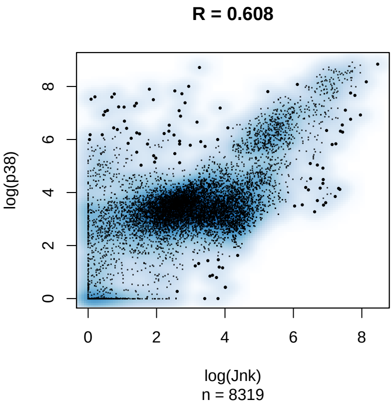
- Two variables are independent conditional on their common parents.
- Conditioning on PKC and PKA, Jnk and p38 should be independent.

Principle II: Conditional independence.



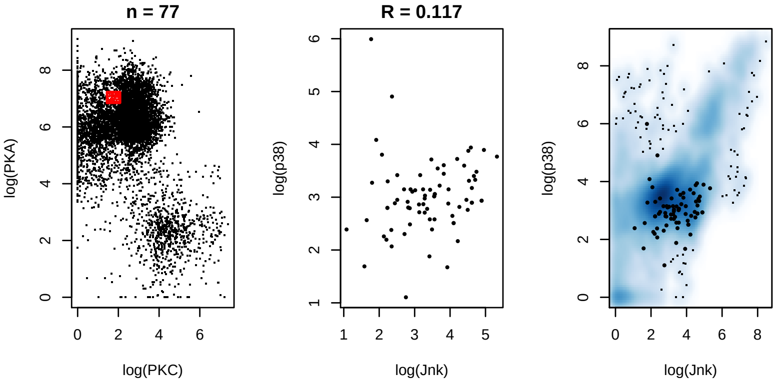
- "Once you and I condition on common factors, we are left with nothing in common."

Principle II: Conditional independence.



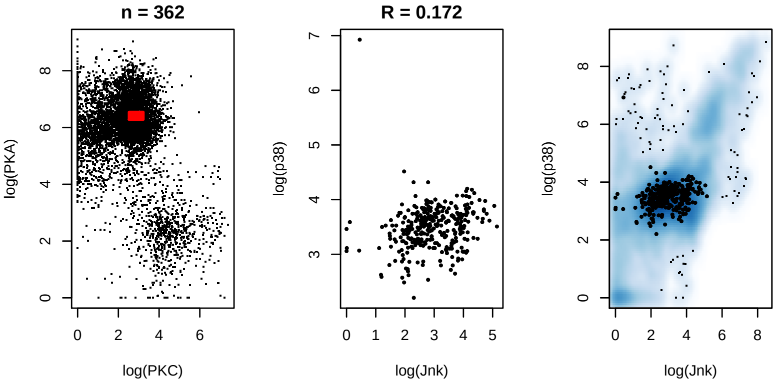
Marginally, p38 and Jnk are correlated.

Principle II: Conditional independence.



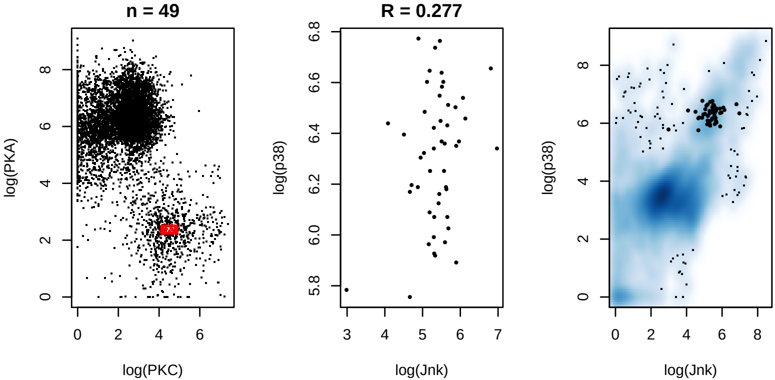
We can't condition on PKA and PKC since the data is continuous. But, conditioning on small windows seems to reduce association.

Principle II: Conditional independence.



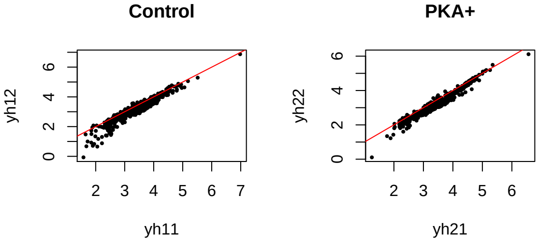
Left: We condition on (PKA, PKC) to lie within the indicated window. *Center:* Conditional joint distribution of (Jnk, p38). *Right:* Conditional join distribution, overlaid on marginal distribution.

Principle II: Conditional independence.



PKA and PKC *explain away* some (if not all) of the association between Jnk and p38. (Recall that R = 0.608 marginally.)

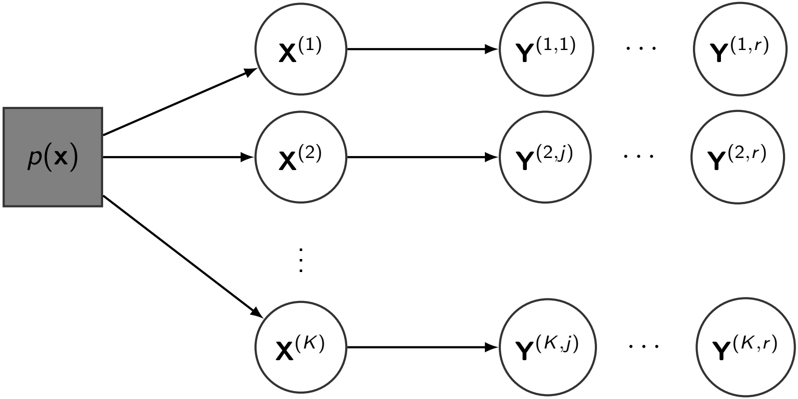
Principle III: Predictive invariance.



A principle brought to attention by a recent paper by the "Zurich group" (Peters Meinshausen, Buhlmann).

Look! A diagram!

Don't put this in the final presentation.



Legend: $K = \{$ 2, 9, 99, 999 $\}$