Taming the challenge of extrapolation:
From multiple experiments and observations to valid causal conclusions

Judea Pearl and Elias Bareinboim, UCLA

One distinct feature of big data applications, setting them apart from traditional causal analyses, is that data are pooled together from multiple sources. Each represents a different population, studied under different conditions, some experimental and some observational, and each contaminated by different confounding factors and different sample selection mechanisms. All of these studies need be pooled together to answer causal questions in yet a new environment, unmatched by any of those studied.

The problem, as described above, appears utterly hopeless, almost as hopeless as confronting physics problems in the 16th century, before the advent of algebra.

My talk will demonstrate that the solutions to the two types of problems, data pooling on the one hand and algebraic equations on the other, are strikingly similar; they can be unraveled by symbolic operations and managed at ease by anyone who masters the two basic tools of causal inference: graphical models and the logic of causation.

The net result is the emergence of algorithms that decide what piece of information must be extracted from each of the available data sources and how to combine those pieces together so as to provide valid answers to a variety of causal questions: What if? How? and Why?

To gain an appreciation for the task, the audience may wish to consider three toy problems described in Figure 3 of this introductory paper http://ftp.cs.ucla.edu/pub/stat_ser/r400.pdf All three problems involve the same task: We gather causal information in one environment and we wish to generalize it to another, which differs from the first in a set Z of characteristics. How do we adjust the available information so as to account for differences between the two populations?

It turns out, as readers will immediately discover, that there is no universal adjustment formula for this task; the adjustment should vary from case to case depending on the location of the set Z in the causal scheme of things. Traditional schemes like post-stratification or re-weighting would not work, except in special cases. Fortunately, however, once the set Z is allocated in the causal graph, the proper adjustment can be generated automatically using symbolic operations, in much the same way that we solve algebraic equations.

The talk will describe how we go from two to multiple environments, how we characterize the idiosyncratic features of each population or environment, and how we can generalize from unrepresentative samples to population level effects,

Related papers can be viewed here:
Tutorials
http://ftp.cs.ucla.edu/pub/stat_ser/r350.pdf
http://ftp.cs.ucla.edu/pub/stat_ser/r424-reprint.pdf
Extrapolation problem
http://ftp.cs.ucla.edu/pub/stat_ser/r425.pdf
http://ftp.cs.ucla.edu/pub/stat_ser/r407.pdf
http://ftp.cs.ucla.edu/pub/stat_ser/r387.pdf

FUSION + EXTRAPOLATION:

FROM MULTIPLE EXPERIMENTS AND OBSERVATIONS TO VALID CAUSAL CONCLUSIONS

Judea Pearl
Elias Bareinboim

University of California, Los Angeles
(www.cs.ucla.edu/~judea/)

THE DATA POOLING PROBLEM

The problem

- How to combine results of several experimental and observational studies, each conducted on a different population and under a different set of conditions,
- so as to construct a valid estimate of effect size in yet a new population, unmatched by any of those studied.

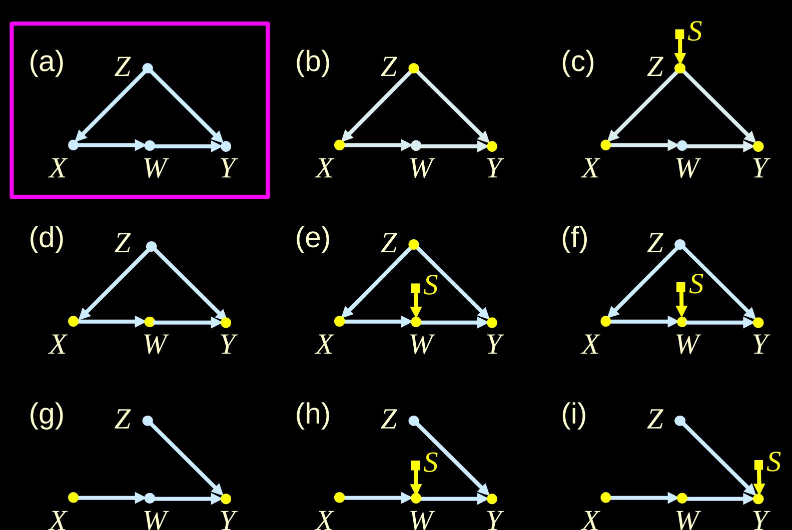
THE PROBLEM IN REAL LIFE

Target population $\prod^*$ Query of interest: $Q = P^*(y \mid do(x))$

(a) Arkansas Survey data available	**(b) New York** Survey data Resembling target	**(c) Los Angeles** Survey data Younger population
(d) Boston Age not recorded Mostly successful lawyers	**(e) San Francisco** High post-treatment blood pressure	**(f) Texas** Mostly Spanish subjects High attrition
(g) Toronto Randomized trial College students	**(h) Utah** RCT, paid volunteers, unemployed	**(i) Wyoming** RCT, young athletes

THE PROBLEM IN MATHEMATICS

Target population $\prod^*$ Query of interest: $Q = P^*(y \mid do(x))$



FIVE LESSONS FROM THE THEATRE OF CAUSAL INFERENCE

1. Every causal inference task must rely on judgmental, extra-data assumptions.
2. We have ways of encoding those assumptions mathematically.
3. We have a mathematical machinery to take those assumptions, combine them with data and derive answers to questions of interest.
4. We have a way of doing (2) and (3) in a language that permits us to judge the scientific plausibility of our assumptions and to derive their ramifications swiftly and transparently.
5. Items (2)-(4) make the fusion problem manageable.

OUTLINE

Three useful tools

- Reading a diagram
- Representing interventions
- *Do*-calculus: evaluating interventions

Transportability: The two-population problem

- Three toy examples and their solutions
- Why transportation requires causal calculus

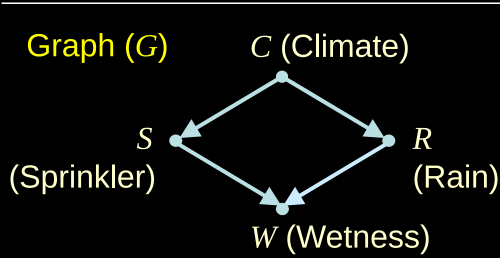
The mathematics of solving extrapolation problems

- The inference engine
- Examples

6

READING A DIAGRAM

Graph (G)



Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

Miracles do happen

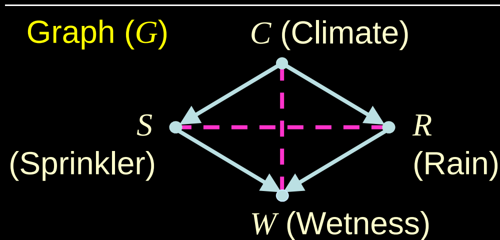
If the U's are independent, the observed distribution $P(C,R,S,W)$ satisfies constraints that are:

(1) independent of the f's and of $P(U)$,

(2) readable from the graph.

READING A DIAGRAM (Cont.)

Graph (G)



Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

Every missing arrow advertises an independency, conditional on a separating set.

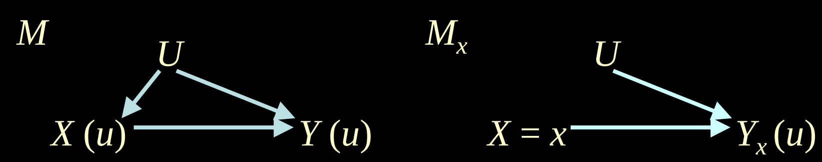
e.g., $C \perp\!\!\!\perp W \mid (S, R)$ $\qquad S \perp\!\!\!\perp R \mid C$

Applications:

1. Model testing
2. Structure learning
3. Reducing "what if I do" questions to symbolic calculus
4. Reducing extrapolation to symbolic calculus

REPRESENTING INTERVENTIONS

Definition:
Given a model M, the effect of setting X to x, $P(Y = y \mid do\,(X{=}x))$, is equal to the probability of $Y = y$ in a mutilated model M_x, in which the equation for X is replaced by $X = x$.



The Fundamental Equation of Interventions:

$$\boxed{P(Y = y \mid do(X = x)) \triangleq P_{M_x}(Y = y)} = P(Y_x = y)$$

THE TWO FUNDAMENTAL LAWS OF CAUSAL INFERENCE

1. The Law of Counterfactuals (and Interventions)

$$\boxed{Y_x(u) = Y_{M_x}(u)}$$

(M generates and evaluates all counterfactuals.)

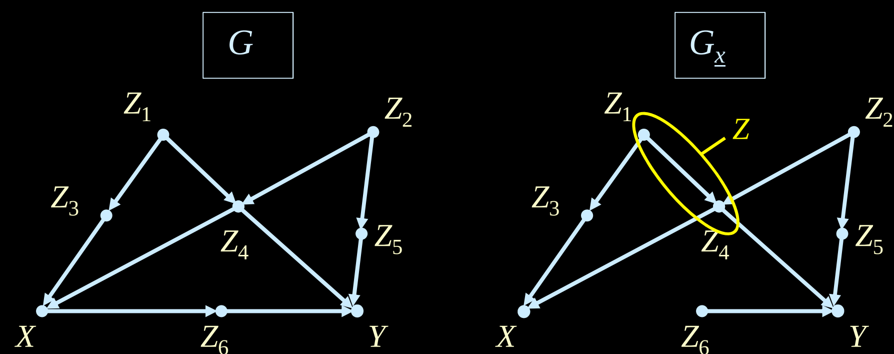
2. The Law of Conditional Independence (d-separation)

$$\boxed{(X \text{ sep } Y \mid Z)_{G(M)} \Rightarrow (X \perp\!\!\!\perp Y \mid Z)_{P(v)}}$$

(Separation in the model ⇒ independence in the distribution.)

ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

$P(y \mid do(x))$ is estimable if there is a set Z of variables that d-separates X from Y in $G_{\underline{x}}$ ("conditional ignorability" $X \perp\!\!\!\perp Y_x \mid Z$)

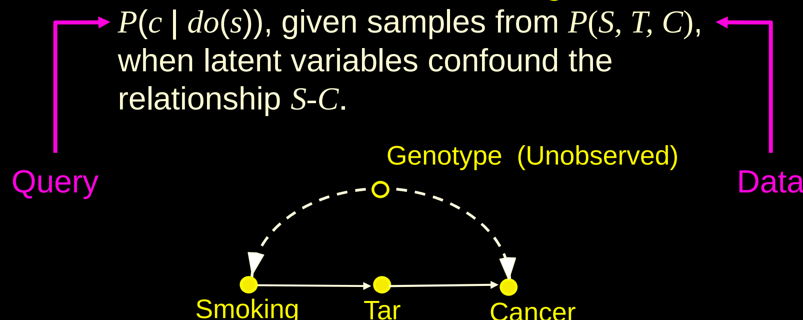


Moreover, $P(y \mid do((x)) = \sum_z P(y \mid x, z) P(z)$

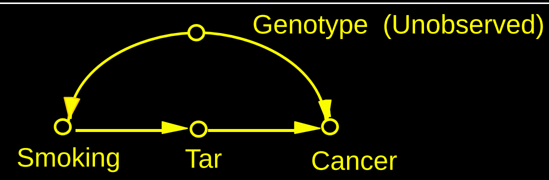
("adjusting" for z)

GOING BEYOND ADJUSTMENT

Goal: Find the effect of *Smoking* on *Cancer*, $P(c \mid do(s))$, given samples from $P(S, T, C)$, when latent variables confound the relationship S-C.



IDENTIFICATION REDUCED TO CALCULUS



Query

$P(c \mid do(s)) = \sum_t P(c \mid do(s), t) P(t \mid do(s))$	Probability Axioms
$= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s))$	Rule 2
$= \sum_t P(c \mid do(s), do(t)) P(t \mid s)$	Rule 2
$= \sum_t P(c \mid do(t) P(t \mid s)$	Rule 3
$= \sum_{s'} \sum_t P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s)$	Probability Axioms
$= \sum_{s'} \sum_t P(c \mid t, s') P(s' \mid do(t)) P(t \mid s)$	Rule 2
$= \sum_{s'} \sum_t P(c \mid t, s') P(s') P(t \mid s)$	Rule 3

Estimand

DO-CALCULUS
(IDENTIFICATION REDUCED TO CALCULUS)

The following transformations are valid for every interventional distribution generated by a structural causal model M:

Rule 1: Ignoring observations

$$P(y \mid do(x), z, w) = P(y \mid do(x), w), \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$$

Rule 2: Action/observation exchange

$$P(y \mid do(x), do(z), w) = P(y \mid do(x), z, w), \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$$

Rule 3: Ignoring actions

$$P(y \mid do(x), do(z), w) = P(y \mid do(x), w), \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\,\overline{Z(W)}}}$$

SUMMARY: THE IDENTIFICATION PROBLEM IS SOLVED (NONPARAMETRICALLY)

- The estimability of any expression of the form

  $Q = P(y_1, y_2, ..., y_n \mid do(x_1, x_2, ..., x_m), z_1, z_2, ..., z_k)$

  can be decided in polynomial time.

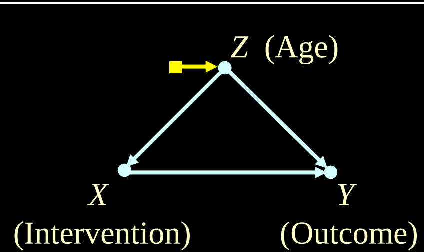
- If Q is estimable, then its estimand can be derived in polynomial time.

- The algorithm is complete.

- Same for ETT (Shpitser 2008).

THE TWO–POPULATION PROBLEM

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT NYC?



$\prod(\text{LA}) \rightarrow \prod{}^*(\text{NY})$

Experimental study in LA

Measured: $P(x, y, z)$

$P(y \mid do(x), z)$

Observational study in NYC

Measured: $P^*(x, y, z)$

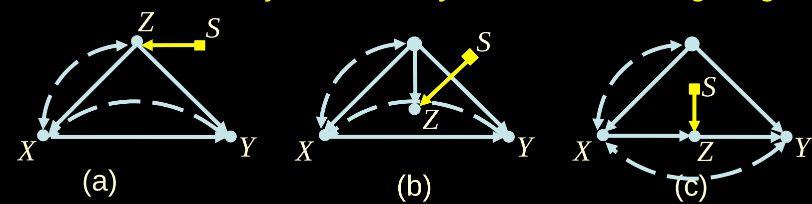
$P^*(z) \neq P(z)$

Needed: $Q = P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$

Transport Formula: $Q = F(P, P_{do}, P^*)$

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

Lesson: Not every dissimilarity deserves re-weighting.



a) ***Z*** represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

b) ***Z*** represents language skill

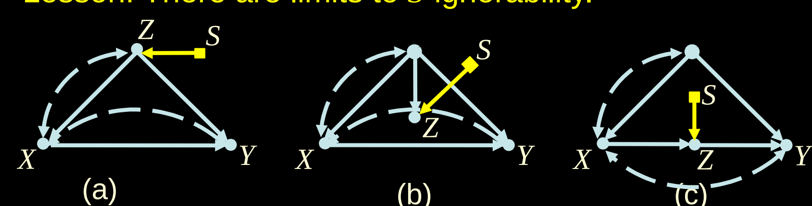
$$P^*(y \mid do(x)) \stackrel{?}{=} P(y \mid do(x))$$

c) ***Z*** represents a bio-marker

$$P^*(y \mid do(x)) \stackrel{?}{=} P(y \mid do(x), z) P^*(z \mid x)$$

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

Lesson: There are limits to *S*-ignorability.



a) ***Z*** represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

b) ***Z*** represents language skill

$$P^*(y \mid do(x)) = P(y \mid do(x))$$

c) ***Z*** represents a bio-marker

$$P^*(y \mid do(x)) = P(y \mid do(x), z) P^*(z \mid x)$$

TRANSPORTABILITY REDUCED TO CALCULUS

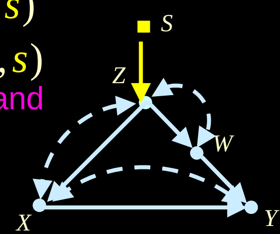
Theorem

A causal relation ***R*** is transportable from $\prod$ to $\prod^*$ if and only if it is reducible, using the rules of ***do***-calculus, to an expression in which ***S*** is separated from ***do***( ).

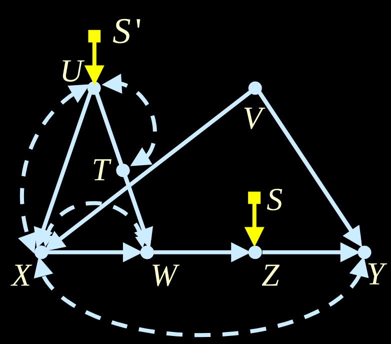
Query

$$R\left(\prod^*\right) = P^*(y \mid do(x)) = P(y \mid do(x), s)$$
$$= \sum_w P(y \mid do(x), s, w) P(w \mid do(x), s)$$
$$= \sum_w P(y \mid do(x), w) P(w \mid s)$$
$$= \sum_w P(y \mid do(x), w) P^*(w)$$

Estimand



RESULT: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



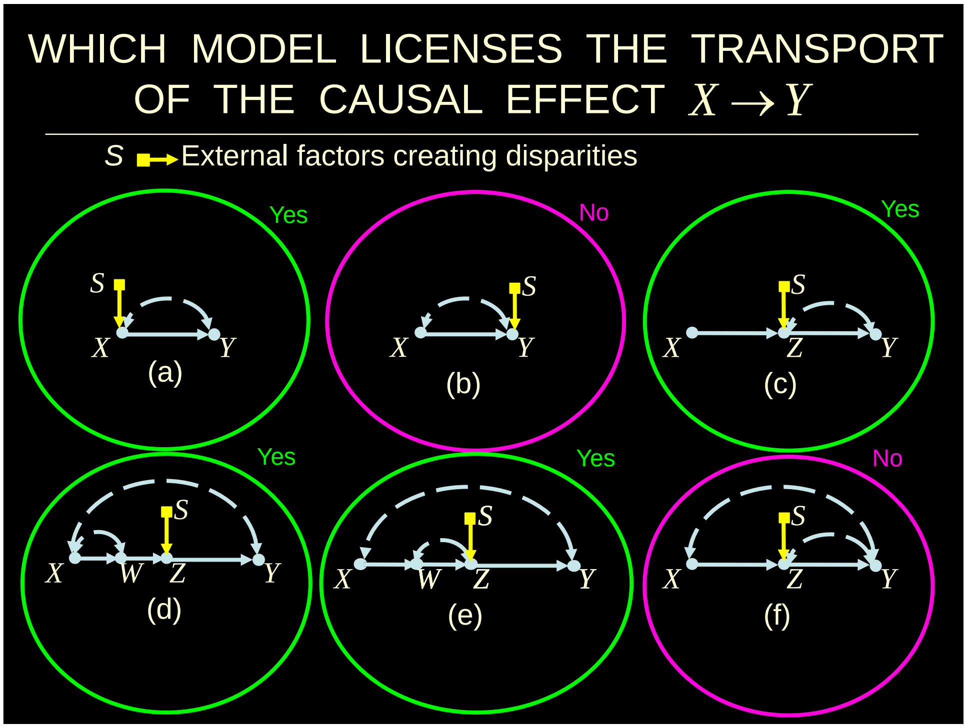
INPUT: Annotated Causal Graph

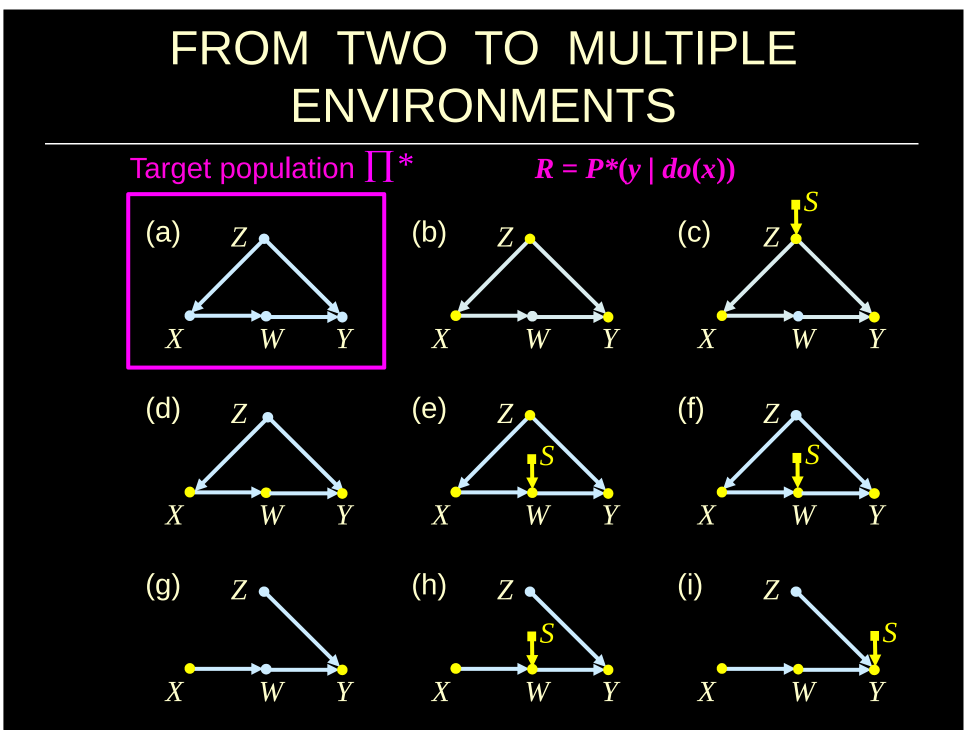
S ➡ Factors creating differences

OUTPUT:

1. Transportable or not?
2. Measurements to be taken in the experimental study
3. Measurements to be taken in the target population
4. A transport formula
5. Completeness (Bareinboim, 2012)

$$P^*(y \mid do(x)) =$$
$$\sum_z P(y \mid do(x), z) \sum_w P^*(z \mid w) \sum_t P(w \mid do(w), t) P^*(t)$$


WHICH MODEL LICENSES THE TRANSPORT OF THE CAUSAL EFFECT $X \rightarrow Y$
S External factors creating disparities
Yes
S
X
Y
(a)
No
S
X
Y
(b)
Yes
S
X
Z
Y
(c)
Yes
S
X
W
Z
Y
(d)
Yes
S
X
W
Z
Y
(e)
No
S
X
Z
Y
(f)

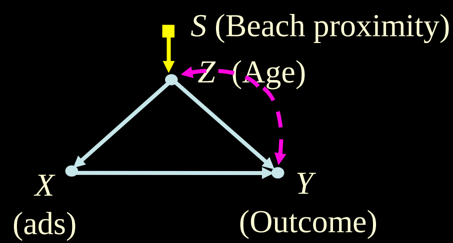

FROM TWO TO MULTIPLE ENVIRONMENTS
Target population $\prod^*$
$R = P^*(y \mid do(x))$
(a) Z X W Y
(b) Z X W Y
(c) S Z X W Y
(d) Z X W Y
(e) Z S X W Y
(f) Z S X W Y
(g) Z X W Y
(h) Z S X W Y
(i) Z S X W Y

SUMMARY OF TRANSPORTABILITY RESULTS

- Nonparametric transportability of experimental results from multiple environments can be determined provided that commonalities and differences are encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The algorithm is complete.

TRANSPORTABILITY VERSUS SAMPLING SELECTION BIAS

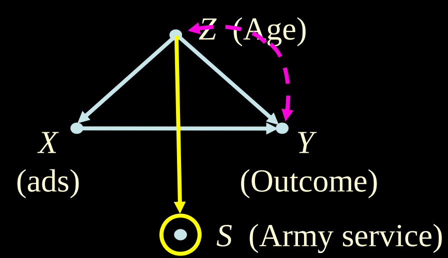
Transportability



S = disparity-producing factors

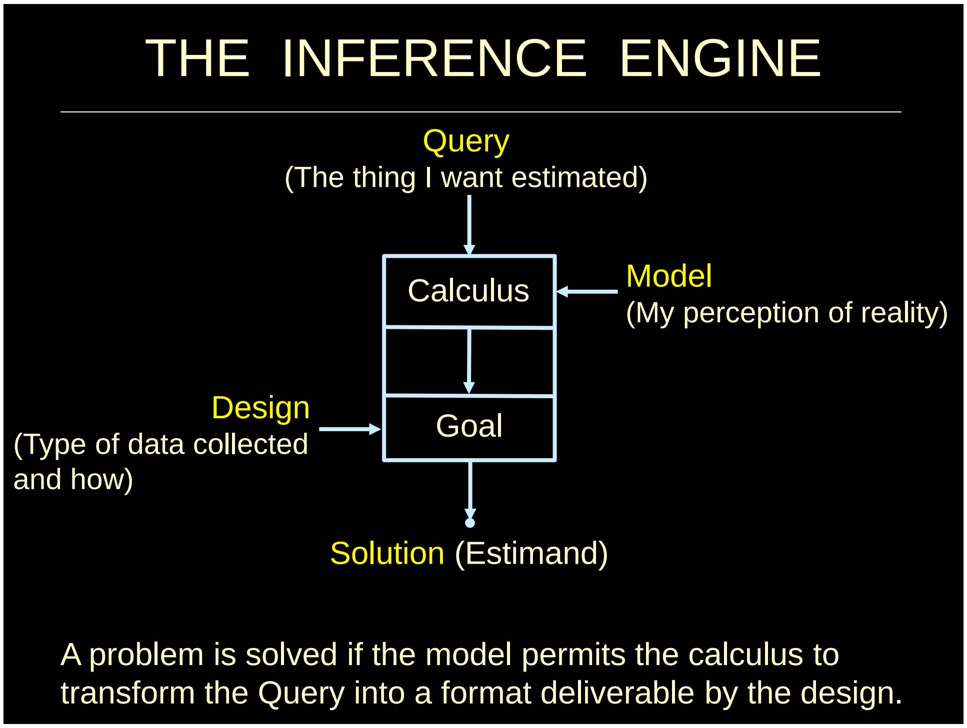
$$P^*(y \mid do(x)) = P(y \mid do(x), S=1)$$
$$= \sum_z P(y \mid do(x), z) P(z \mid S=1)$$

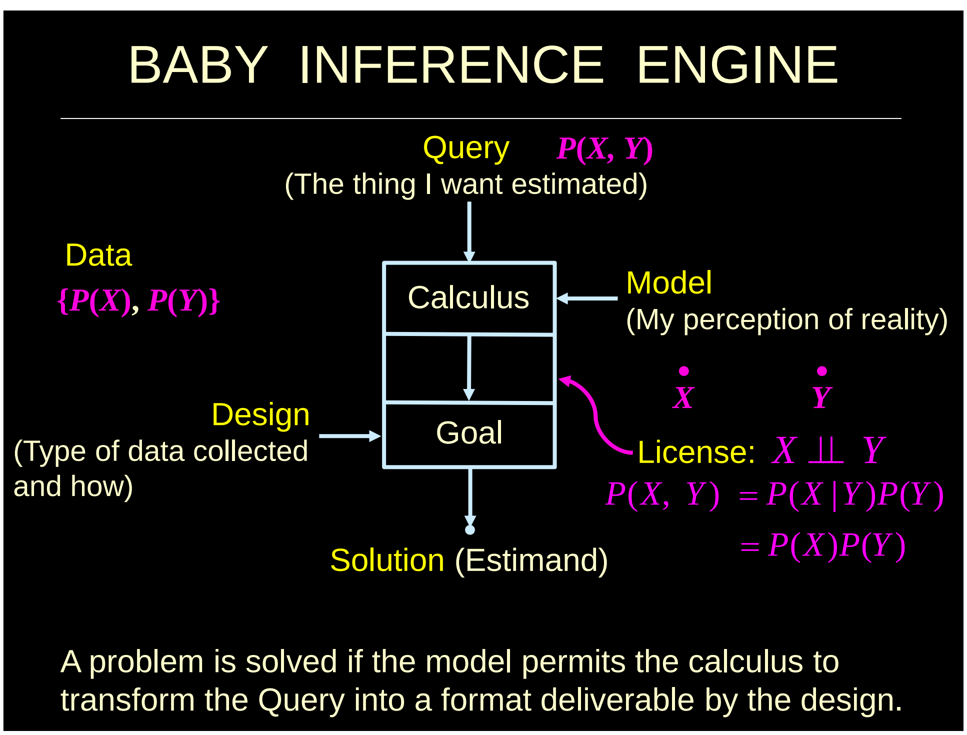
Selection Bias



S = sampling mechanism

$$P^*(y \mid do(x)) = P(y \mid do(x))$$
$$= \sum_z P(y \mid do(x), z, S=1) P(z)$$


THE INFERENCE ENGINE
Query
(The thing I want estimated)
Calculus
Model
(My perception of reality)
Design
(Type of data collected and how)
Goal
Solution (Estimand)
A problem is solved if the model permits the calculus to transform the Query into a format deliverable by the design.


BABY INFERENCE ENGINE
Query $P(X, Y)$
(The thing I want estimated)
Data
$\{P(X), P(Y)\}$
Calculus
Model
(My perception of reality)
X
Y
Design
(Type of data collected and how)
Goal
License: $X \perp\!\!\!\perp Y$
$P(X, Y) = P(X|Y)P(Y)$
$= P(X)P(Y)$
Solution (Estimand)
A problem is solved if the model permits the calculus to transform the Query into a format deliverable by the design.

JUVENILE INFERENCE ENGINE

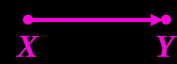
Query $P(y \mid do(x))$
(The thing I want estimated)

Data
$P(X, Y)$

Calculus

Goal

Model
(My perception of reality)



License: No back-door

Design
(Type of data collected and how)

Solution (Estimand) $P(Y \mid X)$

A problem is solved if the model permits the calculus to transform the Query into a format deliverable by the design.

THE TRANSPORT ENGINE

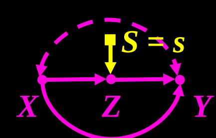
Query $P^*(y \mid do(x)) = P(y \mid do(x), S = s)$
(The thing I want estimated)

Data
$P(y \mid do(x), Z)$
$P(X, Y, Z \mid S = s)$

Calculus

Goal

Model
(My perception of reality)



$\sum_z P(y \mid do(x), z) P(z \mid x, s)$

Design
(Type of data collected and how)

Solution (Estimand)

A problem is solved if the model permits the calculus to transform the Query into a format deliverable by the design.

THE SELECTION-BIAS ENGINE

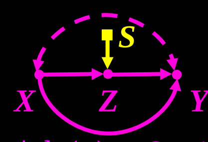
Query $P^*(y \mid do(x)) = P(y \mid do(x))$
(The thing I want estimated)

Data
$P(y \mid do(x), Z, S = 1)$
$P(X, Y, Z)$

Calculus

Goal

Model
(My perception of reality)



$$\sum_z P(y \mid do(x), z, S = 1) P(z \mid x)$$

Design
(Type of data collected and how)

Solution (Estimand)

A problem is solved if the model permits the calculus to transform the Query into a format deliverable by the design.

FIVE LESSONS FROM THE THEATRE OF CAUSAL INFERENCE

1. Every causal inference task must rely on judgmental, extra-data assumptions.
2. We have ways of encoding those assumptions mathematically.
3. We have a mathematical machinery to take those assumptions, combine them with data and derive answers to questions of interest.
4. We have a way of doing (2) and (3) in a language that permits us to judge the scientific plausibility of our assumptions and to derive their ramifications swiftly and transparently.
5. Items (2)-(4) make causal inference easy, fun, and extremely productive.

CONCLUSIONS

"More has been learned about causal inference in the last few decades than the sum total of everything that had been learned about it in all prior recorded history."

(Gary King, Harvard, 2014)

More will be learned about causal inference in the next decade than most of us imagine today.

Thank you

More:
http://bayes.cs.ucla.edu/csl_papers.html