Exchangeability and Invariance

Jiji Zhang
Division of Humanities and Social Sciences
California Institute of Technology

Outline

- Motivation
- Spirtes-Glymour-Scheines's theory of invariance under intervention
- A parallel theory of exchangeability

Simpson's Reversal

Total

	recover	not recover
treatment	20	20
control	16	24

P(recover | treatment) > P(recover | control)

Male

	recover	not recover
treatment	18	12
control	7	3

Female

	recover	not recover
treatment	2	8
control	9	21

P(recover | treatment) < P(recover | control)

Paradoxical Reading

Don’t tell me your sex.
Take it! It is good for people

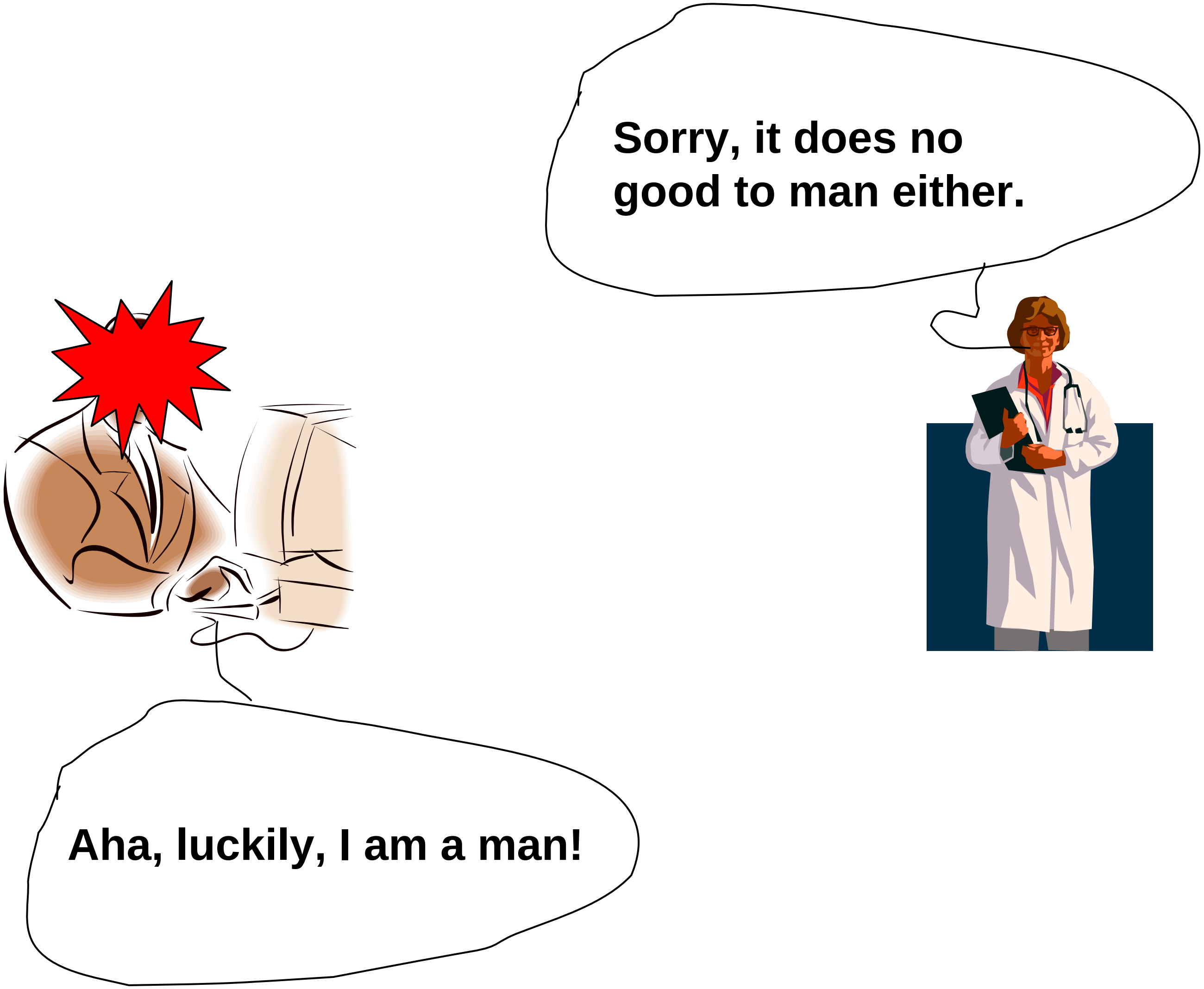
Hi, doctor.
Shall I take the treatment?

Paradoxical Reading

Then you wouldn’t want to take it. It does no good to woman.

What??
What if I were a woman?

Paradoxical Reading



Lindley-Novick's Account

- A new patient of unknown sex should not be given the treatment.

- Despite the frequency distribution, the (personal) probability concerning the new patient satisfies the following:

$$p\ (\text{recover} \mid \text{treatment}) < p\ (\text{recovery} \mid \text{control})$$

A Further Twist

Total

	high	low
white	20	20
black	16	24

P(high yield | white) > P(high yield | black)

Tall

	high	low
white	18	12
black	7	3

Short

	high	low
white	2	8
black	9	21

P(high yield | white) < P(high yield | black)

Different Advice

- In the agricultural case, for the sake of high yield, a new plant should be chosen from the white variety.

- According to Lindley-Novick, the (personal) probability concerning the new plant satisfies the following:

$$p\ (\text{high yield} \mid \text{white}) > p\ (\text{high yield} \mid \text{black})$$

Exchangeability

- **Definition 1 (Exchangeability)** units <$u_1, \ldots, u_m$> are (judged to be) exchangeable in **X** if for every permutation π of the units,

$$p(\mathbf{X}_1, \ldots, \mathbf{X}_m) = p(\mathbf{X}_{\pi(1)}, \ldots, \mathbf{X}_{\pi(m)})$$

- **Definition 2 (Conditional Exchangeability)** units <$u_1, \ldots, u_m$> are (judged to be) exchangeable in **X** conditional on **Y** if for every value setting **y** of **Y** and every permutation π of the units,

$$p(\mathbf{X}_1, \ldots, \mathbf{X}_m \mid \mathbf{Y}_i = \mathbf{y}, \text{ for all } i) = p(\mathbf{X}_{\pi(1)}, \ldots, \mathbf{X}_{\pi(m)} \mid \mathbf{Y}_i = \mathbf{y}, \text{ for all } i)$$

Lindley-Novick's Analysis

- In the medical case, the new unit is exchangeable with the previous units in "recovery" (Y) conditional on "sex" (Z) and "treatment" (X), and exchangeable in "sex" (Z). Hence

$$p(Y \mid X, Z) = P(Y \mid X, Z), \qquad p(Z) = P(Z)$$

  Moreover, $p(Z \mid X) = p(Z)$, from which we can derive that

$$p(Y{=}1 \mid X{=}1) = 0.4 < p(Y{=}1 \mid X{=}0) = 0.5$$

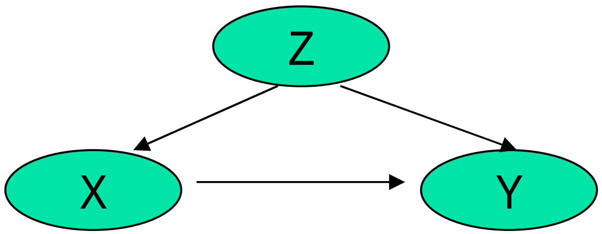
- In the agricultural case, the new unit is exchangeable with the previous units in "yield" (Y) conditional on "variety" (X). Hence

$$p(Y{=}1 \mid X{=}1) = P(Y{=}1 \mid X{=}1) = 0.5$$
$$> p(Y{=}1 \mid X{=}0) = P(Y{=}1 \mid X{=}0) = 0.4$$

Meek-Glymour-Pearl’s Analysis

- Distinguish between *conditioning* and *intervening*.
- From the data we can estimate conditional probabilities, but they are different from probabilities resulting from interventions.
- Causal mechanism matters in calculating probabilities resulting from interventions. The crucial difference between the two cases lies in the causal mechanism that generated the data.

Medical Case



$P_{X:=1}(Y=1)$

$= P_{X:=1}(Y=1 \mid X=1)$

$= P_{X:=1}(Y=1 \mid X=1, Z=1)\, P_{X:=1}(Z=1 \mid X=1)$
$\quad + P_{X:=1}(Y=1 \mid X=1, Z=0)\, P_{X:=1}(Z=0 \mid X=1)$

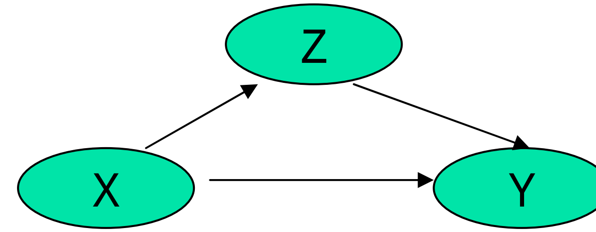
$= P_{X:=1}(Y=1 \mid X=1, Z=1)\, P_{X:=1}(Z=1)$
$\quad + P_{X:=1}(Y=1 \mid X=1, Z=0)\, P_{X:=1}(Z=0)$

$= P(Y=1 \mid X=1, Z=1)\, P(Z=1)$
$\quad + P(Y=1 \mid X=1, Z=0)\, P(Z=0)$

$= 0.4$

Likewise, $P_{X:=0}(Y=1) = 0.5$

Agricultural Case



$P_{X:=1}(Y=1)$

$= P_{X:=1}(Y=1 \mid X=1)$

$= P(Y=1 \mid X=1) = 0.5$

$P_{X:=0}(Y=1)$

$= P_{X:=0}(Y=1 \mid X=0)$

$= P(Y=1 \mid X=0) = 0.4$

“The terminology [of exchangeability] seems to us only to suggest that judgments of exchangeability are related, in a way that remains to be clarified, to judgments about uniformity of causal structure, and that an explicit account of the interaction of causal beliefs and probabilities is necessary to understand when exchangeability should be assumed.”

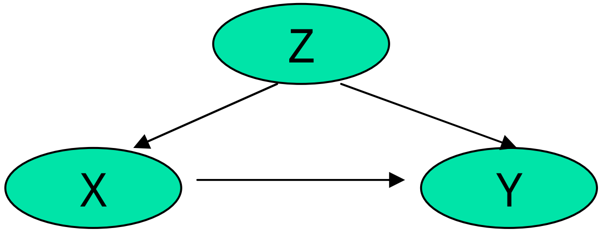
- Meek and Glymour (1994)

“Meek and Glymour keenly observed that the only comprehensible part of Lindley and Novick’s discussion of exchangeability is the one based on causal considerations …

And once causal mechanisms are considered, separate judgment of exchangeability is not needed.”

- Pearl (2000)

Analogy



$P_{X:=1}(Y=1)$

$= P_{X:=1}(Y=1 \mid X=1)$

$= P_{X:=1}(Y=1 \mid X=1, Z=1)\ P_{X:=1}(Z=1 \mid X=1)$
$\quad + P_{X:=1}(Y=1 \mid X=1, Z=0)\ P_{X:=1}(Z=0 \mid X=1)$

$= P_{X:=1}(Y=1 \mid X=1, Z=1)\ P_{X:=1}(Z=1)$
$\quad + P_{X:=1}(Y=1 \mid X=1, Z=0)\ P_{X:=1}(Z=0)$

$= P(Y=1 \mid X=1, Z=1)\ P(Z=1)$
$\quad + P(Y=1 \mid X=1, Z=0)\ P(Z=0)$

$= 0.4$

$p\ (Y=1 \mid X=1)$

$= p(Y=1 \mid X=1, Z=1)\ p(Z=1 \mid X=1)$
$\quad + p(Y=1 \mid X=1, Z=0)\ p(Z=0 \mid X=1)$

$= p(Y=1 \mid X=1, Z=1)\ p(Z=1)$
$\quad + p(Y=1 \mid X=1, Z=0)\ p(Z=0)$

$= P(Y=1 \mid X=1, Z=1)\ P(Z=1)$
$\quad + P(Y=1 \mid X=1, Z=0)\ P(Z=0)$

$=0.4$

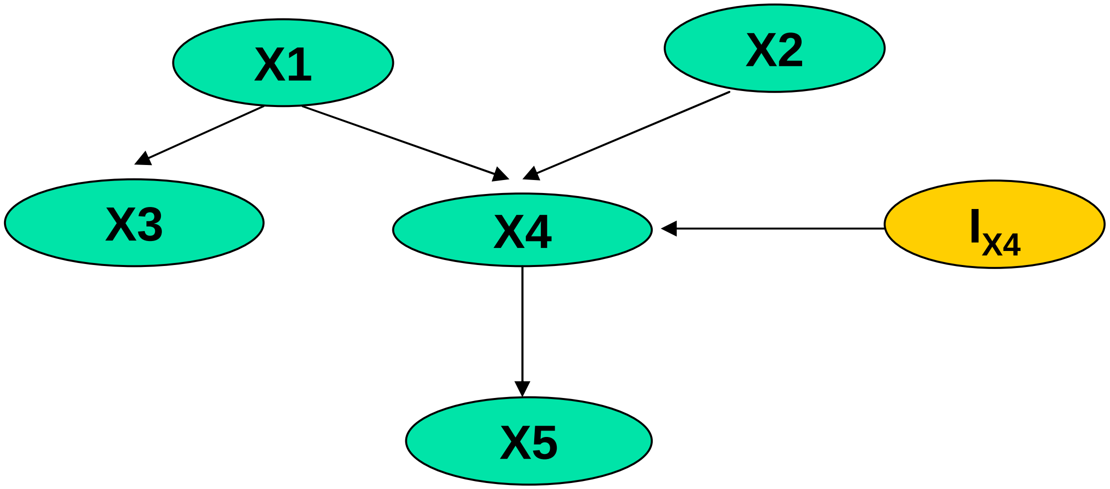
Outline

- Motivation
- Spirtes-Glymour-Scheines’s theory of invariance under intervention
- A parallel theory of exchangeability

General Setup

- Use directed acyclic graphs (DAGs) to represent the causal structure over a (causally sufficient) set of variables of a population.

- Model interventions as exogenous, direct causes.



Features of Interventions

- *Effective*: an intervention on *X* specifies a new conditional distribution of *X* given a (possibly new) set of parents.

- *Local*: an intervention on *X* does not affect the mechanisms that govern other variables in the system.

* *Conservative*: the new set of parents of *X* is a subset of the original set of parents.

Intervention Principle

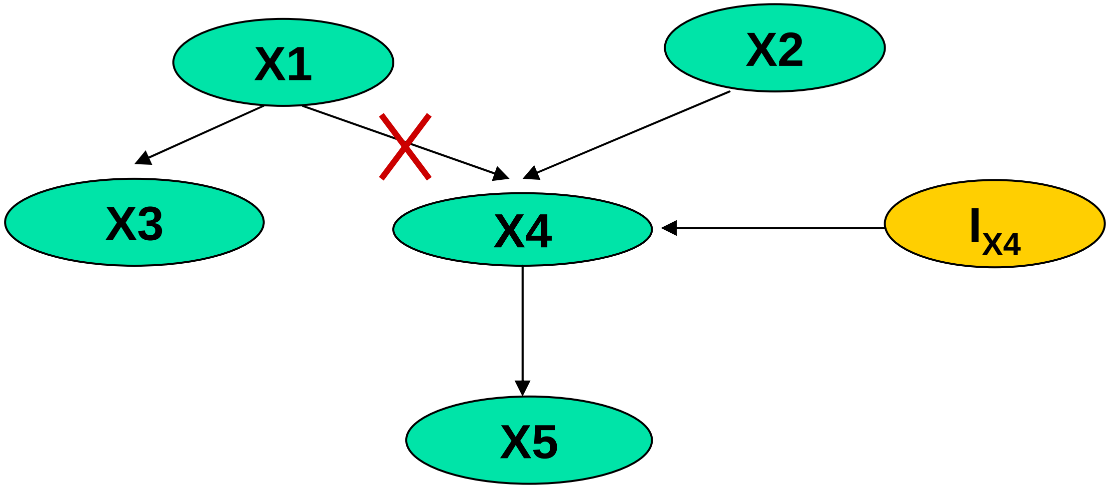
- Let **V** be a causally sufficient system, the causal structure of which is represented by a DAG *G* over **V**. For $\mathbf{X} \subseteq \boldsymbol{V}$, an intervention on **X** may modify the causal structure, represented by *G**.

- *Intervention Principle*: The post-intervention joint distribution of **V** resulting from an intervention on **X** is given by

$$P_{post}(\mathbf{V}) = \prod_{X \in \mathbf{X}} P_{post}(X \mid \mathbf{PA}_{G*}(X)) \prod_{Y \in \mathbf{V} \setminus \mathbf{X}} P_{pre}(Y \mid \mathbf{PA}_{G}(Y))$$

- Intervention principle generalizes the Causal Markov Condition. It assumes that the post-intervention distribution is Markov to the modified causal structure *G**, (and hence also to *G* when *G** is a subgraph of *G*).

Example



$$P_{post}(X1, X2, X3, X4, X5) = P_{post}(X1)P_{post}(X2)P_{post}(X3|X1)\ P_{post}(X4|X2)P_{post}(X5|X4)$$

$$= P_{post}(X4|X2)P_{pre}(X1)P_{pre}(X2)P_{pre}(X3|X1)P_{pre}(X5|X4)$$

Invariance

- For disjoint **Y, Z** $\subseteq$ **V**, the distribution of **Y** conditional on **Z** is *invariant* under an intervention of **X**, if

$$P_{post}(\mathbf{Y} \mid \mathbf{Z}) = P_{pre}(\mathbf{Y} \mid \mathbf{Z})$$

- The intervention principle assumes the following: for every $Y \notin \mathbf{X}$, the distribution of Y conditional on $\mathbf{PA}_G(Y)$ is invariant under an intervention of **X**.

- These basic invariance statement imply others, under the generalized Causal Markov Condition.

D-separation

- In a DAG *G*, *V* is a *collider* on a path *p* if the two edges incident to *V* on *p* are both into *V* ($\rightarrow V \leftarrow$). Otherwise it is a non-collider on *p*.

- In a DAG *G*, a path between *X* and *Y* is *active* given **Z** iff.

  (i) every non-collider on the path is not in **Z**;

  (ii) every collider on the path has a descendant in **Z**.

- **X** and **Y** are *d-separated* by **Z** if *no* path between a variable in **X** and a variable in **Y** is active given *Z*. Otherwise they are *d-connected* by *Z.*

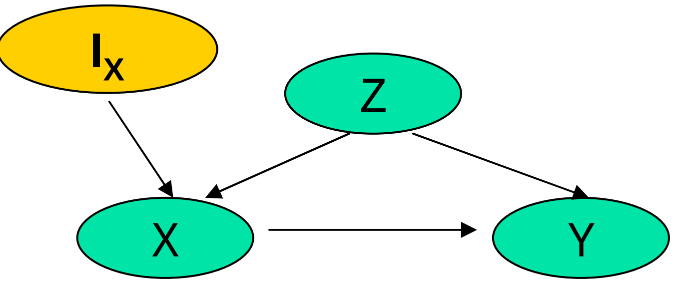
SGS Theorem

- An intervention on *X* can be modeled by adding a *policy variable* to the graph, which is simply an (extra) parent of *X* but otherwise not adjacent to any other variables in *G*.

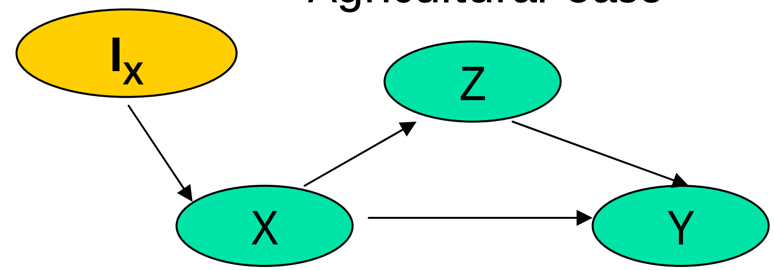
- For $\mathbf{X} \subseteq \mathbf{V}$, the **X**-policy-augmented DAG of *G* is the supergraph of *G* resulting from adding a policy variable for each $X \in \mathbf{X}$.

- **Proposition** (Spirtes, Glymour, Scheines) Let G be the causal DAG for **V**, and consider an intervention on $\mathbf{X} \subseteq \mathbf{V}$. For any two disjoint subsets $\mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, the probability distribution of **Y** conditional on **Z** is invariant under the intervention if in the **X**-Policy-Augmented DAG of G, the policy variables are d-separated from **Y** by **Z**.

Medical Case



Agricultural Case



$$
\begin{aligned}
& P_{X:=1}(Y=1) \\
&= P_{X:=1}(Y=1 \mid X=1) \\
&= P_{X:=1}(Y=1 \mid X=1, Z=1)\, P_{X:=1}(Z=1 \mid X=1) \\
&\quad + P_{X:=1}(Y=1 \mid X=1, Z=0)\, P_{X:=1}(Z=0 \mid X=1) \\
&= P_{X:=1}(Y=1 \mid X=1, Z=1)\, P_{X:=1}(Z=1) \\
&\quad + P_{X:=1}(Y=1 \mid X=1, Z=0)\, P_{X:=1}(Z=0) \\
&= P(Y=1 \mid X=1, Z=1)\, P(Z=1) \\
&\quad + P(Y=1 \mid X=1, Z=0)\, P(Z=0) \\
&= 0.4
\end{aligned}
$$

Likewise, $P_{X:=0}(Y=1) = 0.5$

$$
\begin{aligned}
& P_{X:=1}(Y=1) \\
&= P_{X:=1}(Y=1 \mid X=1) \\
&= P(Y=1 \mid X=1) = 0.5
\end{aligned}
$$

$$
\begin{aligned}
& P_{X:=0}(Y=1) \\
&= P_{X:=0}(Y=1 \mid X=0) \\
&= P(Y=1 \mid X=0) = 0.4
\end{aligned}
$$

Outline

- Motivation
- Spirtes-Glymour-Scheines's theory of invariance under intervention
- A parallel theory of exchangeability

General Setup

- Suppose we know (or believe) that the set of sampled units were governed by a common causal structure over a set of attributes **V**, and hence were exchangeable in **V**.

- A similar, new unit comes in, whose symmetry with the previous units is known to break down (locally) in some ways. Refer to the content of such knowledge or belief as *symmetry-breaking information*.

- What are plausible exchangeability judgments given the information?

Strong Conditional Exchangeability

- Alternative definition of conditional exchangeability:

  <$u_1, \ldots, u_m$> are (judged to be) exchangeable in **X** conditional on **Y** if for every permutation π of the units, and for every possible value settings <$\mathbf{x}_1, \ldots, \mathbf{x}_m$> and <$\mathbf{y}_1, \ldots, \mathbf{y}_m$>,

$$p(\mathbf{X}_1=\mathbf{x}_1,\ldots, \mathbf{X}_m=\mathbf{x}_m \mid \mathbf{Y}_1=\mathbf{y}_1,\ldots, \mathbf{Y}_m=\mathbf{y}_m)$$

$$= p(\mathbf{X}_{\pi(1)}=\mathbf{x}_1,\ldots, \mathbf{X}_{\pi(m)}=\mathbf{x}_m \mid \mathbf{Y}_{\pi(1)}=\mathbf{y}_1,\ldots, \mathbf{Y}_{\pi(m)}=\mathbf{y}_m)$$

- We get the earlier definition by requiring that $\mathbf{y}_1 = \ldots = \mathbf{y}_m$.

- Exchangeability in **Y** and exchangeability in **X** given **Y** in this stronger sense implies exchangeability in (**X**, **Y**).

Default Knowledge Situation

- When there is no symmetry-breaking information, it is natural and plausible to judge that the new unit is exchangeable in **V** with the previous units.

- Suppose the common causal structure for units $<u_1, \ldots, u_m>$ is known to be represented by a DAG *G* over **V**. Here is a subjective, multiple-unit version of the Causal Markov Condition:

$$p(\mathbf{V}_1, \ldots, \mathbf{V}_m) = \prod_{X \in \mathbf{V}} p(X_1, \ldots, X_m \mid \mathbf{PA}_G(X_1), \ldots, \mathbf{PA}_G(X_m))$$

- Given the condition, exchangeability in **V** is equivalent to saying that for every $X \in \mathbf{V}$, the new unit and the previous units are *strongly* exchangeable in *X* conditional on *X*'s causal parents $\mathbf{PA}_G(X)$.

Analogous Postulates

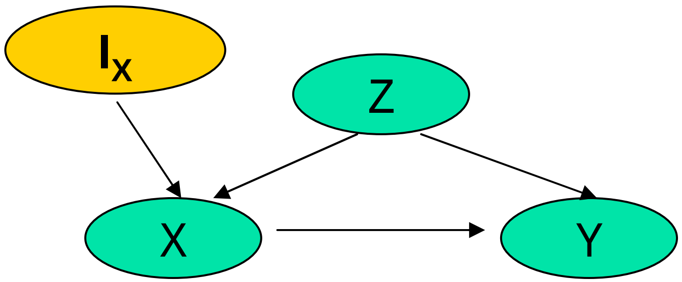
- Local exchangeability judgment: pending symmetry-breaking information concerning $X \in \mathbf{V}$, the new unit is to be judged (strongly) exchangeable with the previous units in X conditional on its causal parents.

- Generalized Markov Condition: the Markov condition still holds relative to G given the symmetry-breaking information that applies to some members of $\mathbf{V}$.

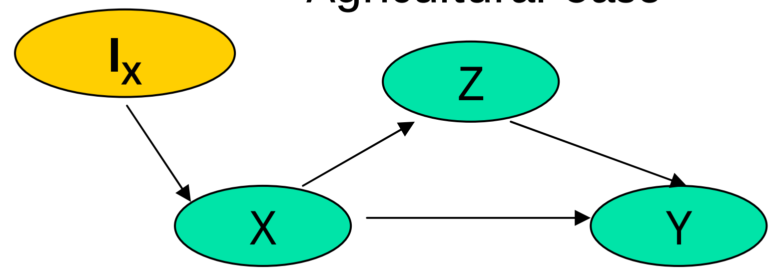
Analogous Result

- We can then exploit dummy variables analogous to SGS' policy variables, which will be called information variables.
- Same with *Information-Augmented DAG* of *G*.
- The local exchangeability judgments and generalized Markov condition imply that *for any two disjoint sets of variables* **X, Y** $\subseteq$ **V**, *if* **X** *is d-separated from the information variables by* **Y** *in the Information-Augmented DAG of G, then the new unit and the previous units should be judged (strongly) exchangeable in* **X** *conditional on* **Y**.

Medical Case



Agricultural Case



Exchangeable in Y given X and Z

Exchangeable in Z

Exchangeable in (Y, Z) given X

Questions about the Postulates

- Does a “causal structure” for a population represent relatively stable exchangeability judgments?

- Is the “multiple-unit” version of the Markov condition more objectionable than the “single-unit” version?