Direct and Indirect Effects
Author: Judea Pearl

Po-Chu Hsu

UC Irvine

2nd March, 2023
1 Introduction

2 Conceptual Analysis
   - Direct versus Total Effects
   - Descriptive versus Prescriptive Interpretation
   - Policy Implications of the Descriptive Interpretation
   - Descriptive Interpretation of Indirect Effects
   - General Path-specific Effects

3 Formal Analysis
   - Notation
   - Controlled Direct Effects
   - Natural Direct Effects: Formulation
   - Natural Direct Effects: Identification
   - Natural Indirect Effects: Formulation
   - Natural Indirect Effects: Identification

4 Conclusions
1. Introduction

Introduction

- **Total Effects:** The probability $Y = y$ when $X$ is set to $x$ by external intervention.

\[ P(Y_x = y) = P(y|do(x)) \]

- **Direct Effects:** The direct effect can be measured by holding all intermediate variables constant.

\[ X \rightarrow Y \]

- **Indirect Effects:** In a linear system, indirect effect $= \text{total effect} - \text{direct effect}$. In nonlinear system, it cannot be measured by holding some variables constant.

- This paper presents a new way of defining the effect transmitted through a restricted set of paths without controlling variables on the remaining paths.

- This permits the assessment of a more natural type of direct and indirect effects, one that is applicable in both linear and nonlinear models and that has broader policy-related interpretations.
2. Conceptual Analysis

2.1. Direct versus Total Effects

Example: Drug, Aspirin and Headache

![Diagram]

- **Total effect:**
  How beneficial the drug is to the population as a whole.

  \[
  P(Y_x = y) - P(Y_{x^*} = y)
  \]

- **Direct effect:**
  Whether aspirin should be encouraged or discouraged during the treatment.

  \[
  P(Y_{xz} = y) - P(Y_{x^*z} = y)
  \]
### Descriptive Versus Prescriptive Interpretation

![Diagram](image)

**Prescriptive:** (controlled effect)
- Whether an untreated patient would improve if treated while holding the aspirin intake fixed at a predetermined level $Z = z$.
- The actual consumption of aspirin under uncontrolled conditions need not concern.
- $X = x^*$ before taking.
- $X = x$ after taking.

$$P(Y_{xz} = y) - P(Y_{x^*z} = y)$$

**Descriptive:** (natural effect)
- Whether the untreated patient would improve if treated, but hold aspirin intake at the level the patient currently consumes under no treatment condition.
- The descriptive formulation requires knowledge of the individual natural behavior.
- Requires testing the same group of patients twice, which is rarely feasible.

$$P(Y_x = y|z) - P(Y_{x^*} = y|z)$$

---

Po-Chu Hsu (UCI)
Policy makers are more interested in **average natural direct effect**.

In the real world, it is impossible to control all patients’ aspirin usage to the same level since aspirin usage is different from individual to individual.

The direct effect of descriptive interpretation carries operational implications and better fits the practical situation.
2. Conceptual Analysis ▷ 2.4. Descriptive Interpretation of Indirect Effects

Descriptive Interpretation of Indirect Effects

- The descriptive conception of direct effects can be transported to the formulation of indirect effects.

- For example, to assess the natural indirect effect of a specific patient, we withhold treatment and ask whether that patient would recover if given as much aspirin as he would have taken if he had been under treatment.

- In contrast, the prescriptive formulation is not transportable since there is no way of preventing the direct effect from operating by holding certain variables constant.
General Path-Specific Effects

- Path-specific Effects can be understood by path-deactivation process.
- A selected set of paths are forced to remain inactive during the transition from $X = x^*$ to $X = x$.
- (a) If we evaluate $X \rightarrow Z \rightarrow W \rightarrow Y$, we cannot hold $Z$ or $W$ constant.
- (b) We can isolate the direct effect by replacing $x$ with $x^*$ in the equation for $W$ and replace $z$ with $z^*(u) = Z_{x^*}(u)$ for $Y$.

![Diagram](a) ![Diagram](b)
3. Formal Analysis ▷ 3.1. Notation

**Notation**

- Let $X$ be the control variable (whose effect we seek to assess).
- Let $Y$ be the response variable.
- Let $Z$ stand for the set of all intermediate variables between $X$ and $Y$.
- Use $Y_x(u)$ to denote the value $Y$ in unit $U = u$ under the control of $do(X = x)$. 

![Diagram](image-url)
### 3.2. Controlled Direct Effects

#### Controlled Direct Effects (CDE) (Review)

- Given a causal model $M$ with causal graph $G$, the Controlled Direct Effect (CDE) of $X = x$ on $Y$ in unit $U = u$ and setting $Z = z$ is given by

$$CDE_z(x, x^*; Y, u) = Y_{xz}(u) - Y_{x^*z}(u)$$

- Given a probabilistic causal model $\langle M, P(u) \rangle$, the controlled direct effect of event $X = x$ on $Y$ is defined as:

$$CDE_z(x, x^*; Y) = E(Y_{xz} - Y_{x^*z})$$

where the expectation is taken over $u$. 

Po-Chu Hsu (UCI)
Natural Direct Effects: Formulation

- Given a causal model $M$ with causal graph $G$, the Natural Direct Effect (NDE) of $X = x$ on $Y$ in unit $U = u$ and setting $Z = z$ is given by

$$NDE_z(x, x^*; Y, u) = Y_{xz^*}(u) - Y_{x^*}(u)$$

$Z_{x^*}(u)$ means the $z$ when $X = x^*$.

- Given a probabilistic causal model $\langle M, P(u) \rangle$, the average direct effect of event $X = x$ on $Y$ is defined as:

$$NDE_z(x, x^*; Y) = E(Y_{xz^*}) - E(Y_{x^*})$$

where the expectation is taken over $u$. 
It is difficult to evaluate the average natural direct effect from empirical data.

Formally, this means \( \text{NDE}_z(x, x^*; Y) = E(Y_{xZ_{x^*}}) - E(Y_{x^*}) \) is not reducible to expressions of the form \( P(Y_x = y) \) or \( P(Y_{xz} = y) \).

\( P(Y_x = y) \) governs the causal effect of \( X \) on \( Y \) (obtained by randomizing \( X \)).

\( P(Y_{xz} = y) \) governs the causal effect of \( X \) and \( Z \) on \( Y \) (obtained by randomizing both \( X \) and \( Z \)).

Following pages shows the reduction is feasible.
3. Formal Analysis ▷ 3.4. Natural Direct Effects: Identification

Natural Direct Effects: Experimental Identification

- If there exists a set $W$ of covariates, nondescendants of $X$ or $Z$, such that

$$Y_{xz} \perp \perp Z_{x^*} | W$$

$Y_{xz}$ is conditionally independent of $Z_{x^*}$, given $W$
Natural Direct Effects: Experimental Identification

- If there exists a set $W$ of covariates, nondescendants of $X$ or $Z$, such that
  \[ Y_{xz} \perp Z_{x^*} | W \]
  $Y_{xz}$ is conditionally independent of $Z_{x^*}$, given $W$

- The average natural direct effect is experimentally identifiable by
  \[
  NDE_z(x, x^*; Y) = E(Y_x Z_{x^*}) - E(Y_{x^*})
  \]
  \[
  E(Y_x Z_{x^*}) = \sum_w \sum_z E(Y_{xz} | Z_{x^*} = z, w) P(Z_{x^*} = z | w) P(w)
  \]
  \[
  = \sum_w \sum_z E(Y_{xz} | w) P(Z_{x^*} = z | w) P(w)
  \]
  \[
  E(Y_{x^*}) = E(Y_{x^*} Z_{x^*}) \quad \text{law of composition}
  \]
  \[
  = \sum_w \sum_z E(Y_{x^* z} | w) P(Z_{x^*} = z | w) P(w)
  \]
  \[
  NDE_z(x, x^*; Y) = \sum_{w,z} [E(Y_{xz} | w) - E(Y_{x^* z} | w)] P(Z_{x^*} = z | w) P(w)
  \]
3. Formal Analysis ▷ 3.4. Natural Direct Effects: Identification

Natural Direct Effects: Experimental Identification

(a) A causal model with latent variables ($\mathcal{U}$’s) where the natural direct effect can be identified in experimental studies.

(b) The subgraph $G_{XZ}$ illustrating the criterion of experimental identifiability: $(Y \perp Z \mid W)_{G_{XZ}}$, $W$ d-seperates $Y$ from $Z$ in the graph formed by deleting all (solid) arrows emanating from $X$ and $Z$. 
The identification of the natural direct effect from nonexperimental data requires stronger conditions.

\[ NDE_z(x, x^*; Y) = \sum_{w, z} [E(Y_{xz}|w) - E(Y_{x^*z}|w)]P(Z_{x^*} = z|w)P(w) \]

From the above equation, it is sufficient to identify the conditional probability if

- \( Y_{xz} \perp \perp Z_{x^*}|W \)
- \( P(Y_{xz} = y|W = w) \) is identifiable
- \( P(Z_{x^*} = z|W = w) \) is identifiable
The condition $Y_{xz} \perp Z_{x^*}|W$ still holds when $W = \emptyset$.

Since $Y_{xz}$ is independent of all variables in the model.

In Markovian models we have the following three relationships:

1. $P(Y_{xz} = y) = P(y|x, z)$ since $X \cup Z$ is the set of $Y$'s parents.
2. $P(Z_{x^*} = z) = \sum_s P(z|x^*, s)P(s)$
3. $P(Y_{x, Z_{x^*}} = y) = \sum_s \sum_z P(y|x, z)P(z|x^*, s)P(s)$

where $S$ stands for the parents of $Z$ excluding $X$ or any other set satisfying backdoor criterion.

The average natural direct effect in Markovian models can be identifiable from nonexperimental data

$$NDE_z(x, x^*; Y) = \sum_s \sum_z [E(Y|x, z) - E(Y|x^*, z)]P(z|x^*, s)P(s)$$

We can use simple Markovian model that the effect of $X$ on $Z$ is not confounded, i.e. $P(Z_{x^*} = z) = P(z|x^*)$.

The equation can be simplified as

$$NDE_z(x, x^*; Y) = \sum_s \sum_z [E(Y|x, z) - E(Y|x^*, z)]P(z|x^*)$$
3. Formal Analysis ▷ 3.4. Natural Direct Effects: Identification

Natural Direct Effects: Markovian Models

(a) \( NDE_z(x, x^*; Y) = \sum_s \sum_z [E(Y|x, z) - E(Y|x^*, z)]P(z|x^*, s)P(s) \)

(b) Simple Markovian model (\( X \) on \( Z \) is not confounded)

\[
NDE_z(x, x^*; Y) = \sum_s \sum_z [E(Y|x, z) - E(Y|x^*, z)]P(z|x^*)
\]

this expression can be considered as a weighted average of the controlled direct effect \( E(Y|x, z) - E(Y|x^*, z) \) where the intermediate value \( z \) is chosen according to its distribution under \( x^* \)
The Natural Indirect Effect (NIE) can be defined as

\[ NIE(x, x^*; Y, u) = Y_{x^*, Z_x}(u) - Y_{x^*}(u) \]

The average indirect effect can be defined as:

\[ NIE(x, x^*; Y) = E(Y_{x^*, Z_x}) - E(Y_{x^*}) \]

The Total Effect (TE) is

\[ TE(x, x^*; Y) = E(Y_x) - E(Y_{x^*}) \]

We have

\[ TE(x, x^*; Y) = NIE(x, x^*; Y) + NDE(x, x^*; Y) \]

The total effect on \( Y \) of the transition from \( x^* \) to \( x \) is equal to the difference between the indirect effect associated with this transition and the natural direct effect associated with the reverse transition.
Natural Indirect Effects: Identification

- If there exists a set $W$ of covariates, nondescendants of $X$ or $Z$, such that
  \[ Y_{x^*z} \perp \perp Z_x | W \]
  for all $x$ and $z$, the average indirect effect is experimentally identifiable by
  \[
  NIE(x, x^*; Y) = \sum_{w,z} E(Y_{x^*z} | w)[P(Z_x = z | w) - P(Z_{x^*} = z | w)]P(w)
  \]

- In non-experimental studies, the average indirect effect is identifiable when $E(Y_{x^*z} | w)$, $P(Z_x = z | w)$ and $P(Z_{x^*} = z | w)$ are identifiable.

- In a simple Markovian model, the equation can be simplified to
  \[
  NIE(x, x^*; Y) = \sum_z E(Y | x^*, z)[P(z | x) - P(z | x^*)]
  \]
This paper formulates a new definition of path-specific effects that is based on path switching.

Instead of variable fixing, that extends the interpretation and evaluation of direct and indirect effects to nonlinear models.

It is shown that, in nonparametric models, direct and indirect effects can be estimated consistently from both experimental and nonexperimental data, provided certain conditions hold in the causal diagram.

Markovian models always satisfy these conditions.

Using the new definition, the paper provides an operational interpretation of indirect effects.
Question To The Class

- Please briefly explain why the natural indirect effect is difficult to measure.