Sufficient Covariates and Linear Propensity Analysis

Hui Guo, Philip Dawid

University of Cambridge

AISTATS, Sardinia, 13 May 2010

Introduction

1. Our interests:
   - Causal effect from observational studies
   - Adjusting for covariates
   - Dimension reduction of covariates

Introduction

1. Our interests:
   - Causal effect from observational studies
   - Adjusting for covariates
   - Dimension reduction of covariates

2. Variables
   - T: binary treatment (1: treatment; 0: control)
   - Y: real-valued response
   - X: "covariates" with values in $\mathcal{R}^p$

Introduction

Framework: Decision-Theoretic (Dawid 2002):

- non-stochastic regime indicator F_T:

$$F_T = \emptyset: \quad \textit{observational}$$
$$F_T = 1: \quad \textit{interventional: assigned to treatment}$$
$$F_T = 0: \quad \textit{interventional: assigned to control}$$

Introduction

Framework: Decision-Theoretic (Dawid 2002):

- non-stochastic regime indicator F_T:

$$F_T = \emptyset: \quad \textit{observational}$$
$$F_T = 1: \quad \textit{interventional: assigned to treatment}$$
$$F_T = 0: \quad \textit{interventional: assigned to control}$$

- average causal effect

$$\mathrm{ACE} := \mathsf{E}(Y \mid F_T = 1) - \mathsf{E}(Y \mid F_T = 0),$$

alternatively, we write $\mathrm{ACE} = \mathsf{E}_1(Y) - \mathsf{E}_0(Y)$.

Possible properties

Property 1:

$$X \perp\!\!\!\perp F_T$$

covariate

Possible properties

Property 1:

$$X \perp\!\!\!\perp F_T$$

covariate

Property 2:

$$Y \perp\!\!\!\perp F_T \mid (X, T)$$

sufficient covariate

Possible properties

Property 1:

$$X \perp\!\!\!\perp F_T$$

covariate

Property 2:

$$Y \perp\!\!\!\perp F_T \mid (X, T)$$

sufficient covariate

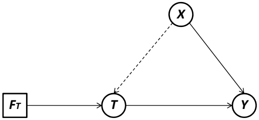
Property 3:

$$P_{\emptyset}(T = t \mid X) > 0$$

strongly sufficient covariate

Possible properties

DAG (influence diagram) for sufficient covariate:



Interplay between various regimes

- specific causal effect:

$$\mathrm{SCE}_X := \mathsf{E}_1(Y \mid X) - \mathsf{E}_0(Y \mid X).$$

Interplay between various regimes

- specific causal effect:

$$\mathrm{SCE}_X := \mathsf{E}_1(Y \mid X) - \mathsf{E}_0(Y \mid X).$$

- **Theorem 2** *For any covariate X, and in each regime $f = \emptyset, 0, 1$,*

$$\mathrm{ACE} = E_f(\mathrm{SCE}_X).$$

Interplay between various regimes

- specific causal effect:

$$\mathrm{SCE}_X := \mathrm{E}_1(Y \mid X) - \mathrm{E}_0(Y \mid X).$$

- **Theorem 2** *For any covariate X, and in each regime $f = \emptyset, 0, 1$,*

$$\mathrm{ACE} = E_f(\mathrm{SCE}_X).$$

- **Theorem 1** *Let X be strongly sufficient. Then for any integrable $Z \preceq (Y, X, T)$, and any versions of the conditional expectations, almost surely in any regime,*

$$E_t(Z \mid X) = E_\emptyset(Z \mid X, T = t), \qquad (t = 0, 1).$$

Reduction of strongly sufficient covariate

- Motivation: easier to adjust for a single rather than multi-dimensional variable, e.g. matching, subclassification.

Reduction of strongly sufficient covariate

- Motivation: easier to adjust for a single rather than multi-dimensional variable, e.g. matching, subclassification.
- If X is strongly sufficient and $V \preceq X$, then V is also strongly sufficient as long as

$$Y \perp\!\!\!\perp F_T \mid (V, T).$$

Reduction of strongly sufficient covariate

- Motivation: easier to adjust for a single rather than multi-dimensional variable, e.g. matching, subclassification.
- If X is strongly sufficient and $V \preceq X$, then V is also strongly sufficient as long as

$$Y \perp\!\!\!\perp F_T \mid (V, T).$$

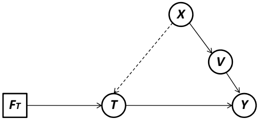
- **Theorem 3** *Suppose X is strongly sufficient and $V \preceq X$. Then V is strongly sufficient if either of the following conditions is satisfied:*

Reduction of strongly sufficient covariate

(*a*) **Response-sufficient reduction:**

$$Y \perp\!\!\!\perp X \mid (V, F_T = t).$$

DAG expression:

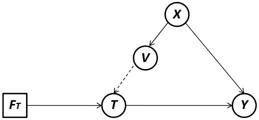


Reduction of strongly sufficient covariate

(b) **Treatment-sufficient reduction:**

$$T \perp\!\!\!\perp X \mid (V, F_T = \emptyset).$$

DAG expression:



Treatment-sufficient reduction

- Let family $\mathcal{Q} = \{Q_0, Q_1\}$ consist of the observational distributions for X, given $T = 0, 1$ *resp.*,

$\rightarrow$ A treatment-sufficient reduction V is a sufficient statistic for $\mathcal{Q}$.

$\rightarrow$ A minimal treatment-sufficient reduction is a minimal sufficient statistic for $\mathcal{Q}$, equivalent to the likelihood ratio statistic $\Lambda := \frac{q_1(X)}{q_0(X)}$.

Propensity variable, propensity score

- Such minimal treatment-sufficient reduction V is termed a *propensity variable* (PV).

- The *propensity score* is

$$\Pi := P_{\emptyset}(T = 1 \mid X) = \frac{\pi \Lambda}{1 - \pi + \pi \Lambda},$$

where $\pi = P_{\emptyset}(T = 1)$.

Example: simple normal linear model

- The joint distribution of (T, X, Y) is given by:

$$P_{\emptyset}(T = 1) = \pi$$

$$X \mid (T, F_T = \emptyset) \sim N(\mu_T, \Sigma)$$

$$Y \mid (X, T, F_T) \sim N(d + \delta T + b'X, \phi),$$

with $\pi \in (0, 1)$, μ_t $(p \times 1)$, Σ (positive definite), b $(p \times 1)$.

Example: simple normal linear model

- The joint distribution of (T, X, Y) is given by:

$$P_{\emptyset}(T = 1) = \pi$$

$$X \mid (T, F_T = \emptyset) \sim N(\mu_T, \Sigma)$$

$$Y \mid (X, T, F_T) \sim N(d + \delta T + b'X, \phi),$$

with $\pi \in (0, 1)$, μ_t $(p \times 1)$, Σ (positive definite), b $(p \times 1)$.

- $X \mid F_T = t$ is taken to be the same as $X \mid F_T = \emptyset$, a normal mixture.
- X is strongly sufficient, we have $\mathrm{ACE} = \delta$ which is also true for any regression on a linear sufficient reduction of X.

Propensity variable *vs.* covariate X

- The log likelihood ratio is

$$\log \Lambda = \mathit{const.} + \mathrm{LD}$$

  with

$$\mathrm{LD} := (\mu_1 - \mu_0)' \Sigma^{-1} X,$$

  where LD is fisher's *linear discriminant* and also a PV in our model.

- **Theorem 4** *The coefficient of T is the same in the linear regressions of Y on (T, LD) as in the linear regression of Y on (T, X).*

Propensity variable *vs.* covariate X

- **Corollary 1** *Given data on* (Y, T, X) *for a sample, Let* LD^* *be the sample linear discriminant. Then the coefficient of* T *in the sample linear regression of* Y *on* T *and* LD^* *is the same as that in the sample linear regression of* Y *on* T *and* X*.*

Propensity variable *vs.* covariate X

- **Corollary 1** *Given data on* (Y, T, X) *for a sample, Let* LD^* *be the sample linear discriminant. Then the coefficient of* T *in the sample linear regression of* Y *on* T *and* LD^* *is the same as that in the sample linear regression of* Y *on* T *and* X.
- Propensity analysis does not increase precision! Adjusting for LD^* rather than for all p predictors makes **absolutely no difference** to our estimate — thus, to its precision.

Simulations

- setting: $X_1 \mid T \sim N(0,1), \qquad X_2 \mid T \sim N(T,1) \quad ind.$

  $$Y \mid (X,T) \sim N(0.5T + X_1, 1)$$

Simulations

- setting: $X_1 \mid T \sim N(0,1), \qquad X_2 \mid T \sim N(T,1) \qquad ind.$

$$Y \mid (X,T) \sim N(0.5T + X_1, 1)$$

- $\mathrm{LP} = X_1$,

  $\mathrm{PV} = \mathrm{LD} = X_2, \qquad Y \mid (X_2, T) \sim N(0.5T, 2).$

Simulations

- setting: $X_1 \mid T \sim N(0,1), \qquad X_2 \mid T \sim N(T,1) \quad ind.$

$$Y \mid (X,T) \sim N(0.5T + X_1, 1)$$

- $\mathrm{LP} = X_1$,

  $\mathrm{PV} = \mathrm{LD} = X_2, \quad Y \mid (X_2, T) \sim N(0.5T, 2).$

- four linear regressions:

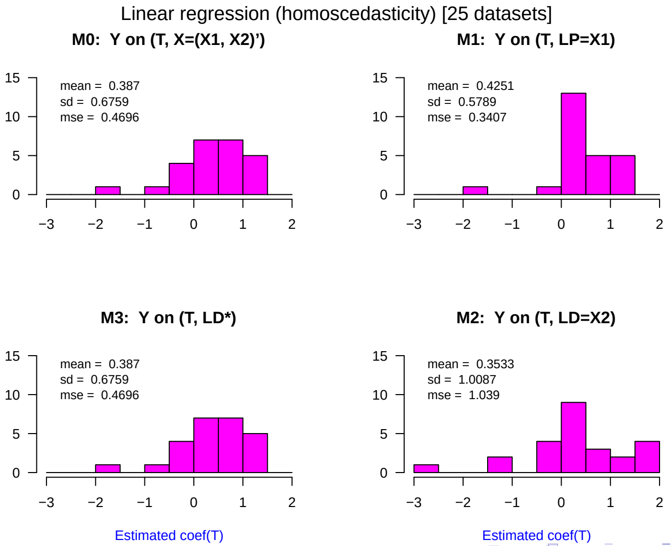
  $M_0 : Y$ on $T, X = (X_1, X_2)'$

  $M_1 : Y$ on $T, \mathrm{LP} = X_1$

  $M_2 : Y$ on $T, \mathrm{PV} = X_2$

  $M_3 : Y$ on T, PV^*.

  All produce unbiased estimator for $\mathrm{ACE} = 0.5$.


Linear regression (homoscedasticity) [25 datasets]
M0: Y on (T, X=(X1, X2)')
mean = 0.387
sd = 0.6759
mse = 0.4696
M1: Y on (T, LP=X1)
mean = 0.4251
sd = 0.5789
mse = 0.3407
M3: Y on (T, LD*)
mean = 0.387
sd = 0.6759
mse = 0.4696
M2: Y on (T, LD=X2)
mean = 0.3533
sd = 1.0087
mse = 1.039
15
10
5
0
−3 −2 −1 0 1 2
Estimated coef(T)
Estimated coef(T)

- A propensity variable is identified as a minimal treatment-sufficient reduction.
- For simple normal linear model, $\mathrm{PV} = \mathrm{LD}$, adjustment for the sample PV^* yields the same estimated ACE as for X. It can neither increase nor decrease precision.
- Our investigations add weight to the accruing evidence that propensity analysis has little to contribute to improving the estimation of causal effects.

THANK YOU !