Ch4. Counterfactuals and Their Applications

Judea Pearl et al.

Presenter : 이종진

Seoul National University

ga0408@snu.ac.kr

January 31, 2020

Table of Contents

Defining and Computing Counterfactuals

Counterfactuals

- Took the road A, 1 hour driving time.
- "I should have taken the road B".
- To emphasize our wish to compare two outcomes under exact same conditions.

Notation

- Using do-expression,

$$E(\text{driving time}|do(\text{Take B}), \text{driving time} = 1)$$

  - leads to clash between the hypothetic driving time and actual driving time.

- Denote hypothetical driving time with subscripts.
- $Y_{X=1}$(or Y_1), $Y_{X=0}$(or Y_0)
- We wish to estimate

$$E(Y_{X=1}|X = 0, Y = Y_0 = 1)$$

Do-expression vs Counterfactual

- Invention queries / retrospective counterfactual queries.
- $E[Y|do(X = x)] = E[Y_x]$
- Population level analysis / Individual level analysis

Defining and computing counterfactuals (in determistic)

- A simple causal model consisting of just three variables, X, Y, U.
- U can be interpreted as individual or situation
- Model M,

$$X = aU$$

$$Y = bX + U$$

- Model M_x,

$$X = x$$

$$Y = bX + U$$

- Compute the counterfactual $Y_x(U)$
- "Y would be y had X been x, in situation U=u"

- a = b = 1, Observe X=2, Y=4. Compute $E(Y_2|X = 2, Y = 4)$
- $4 = 1 * 2 + u,\ Y_2(u) = 1 * 2 + u = 4$

Table 4.1 The values attained by $X(u)$, $Y(u)$, $Y_x(u)$, and $X_y(u)$ in the linear model of Eqs. (4.3) and (4.4)

u	$X(u)$	$Y(u)$	$Y_1(u)$	$Y_2(u)$	$Y_3(u)$	$X_1(u)$	$X_2(u)$	$X_3(u)$
1	1	2	2	3	4	1	1	1
2	2	4	3	4	5	2	2	2
3	3	6	4	5	6	3	3	3

- The do-operator captures the behavior of a population under intervention, whereas $Y_x(u)$ describes the behavior of a specific individual, U=u, under such intervention.

The fundamental Law of Counterfacuals

- $Y_x(u) = Y_{M_x}(u)$
  - $Y_x(u)$ is defined as the solution for Y in the "Surgically modified", submodel M_x
- Consistency rule

$$\text{if } X = x, \text{then } Y_x = y$$

  - $E[Y_x|X = x] = E[Y|X = x]$

The Three Steps in Computing Counterfactuals(determistic)

- Three-step process for computing any deterministic counterfactual:
  1. Abduction: Use evidence E = e to determine the value of U
  2. Action: Modify the model, M, by removing the structural equations for the variables in X and replacing them with the appropriate functions X=x, to obtain the modified model, M_x
  3. Prediction: Use the modified model, M_x, and the value of U to compute the value of Y, the consequence of the counterfactual.

Probability of Counterfactuals

- By assigning probabilities P(U=u) over the exogenous variables U, we can compute probability of counterfactuals.
- In same example, $P(U = 1) = \frac{1}{2}, P(U = 2) = \frac{1}{3}, P(U = 3) = \frac{1}{6}$
- $P(Y_2 = 4) = P(u : Y_2(u) = 3) = \frac{1}{2}$
- $P(Y_2 > 3, Y_1 < 4) = \frac{1}{3}, P(Y_1 < 4, Y - X > 1) = \frac{1}{3}$
- $P(Y_3 > Y)|Y > 2) = \frac{1}{3}/\frac{1}{2} = \frac{2}{3}$
- We can find joint probability of two events occuring in two different words.

The Three Steps in Computing Counterfactuals(nondetermistic)

- Three-step process to any probabilistic nonlinear system.
  1. Abduction: Update P(U) by the evience to obtain P(U|E=e)
  2. Action: Modify the model, M, by removing the structural equations for the variables in X and replacing them with the appropriate functions X=x, to obtain the modified model, M_x
  3. Prediction: Use the modified model, M_x, and the updated probabilities over the U variables, P(U|E=e), to compute the expectation of Y, the consequence of the counterfactual.

Example

$$X = U_1$$

$$Z = aX + U_2$$

$$Y = bZ$$

- X =1: stand for having a college education
- $U_2 = 1$: for having professional experience
- Z: for the level of skill needed for a given job, Y for salary, mediator
- Y: Salary

Example

- $E[Y_{X=x}|Z=1]$ vs $E[Y|do(X=1), Z=1]$
- The former is the expected salary of individual with skill level Z=1, had they received a college education.
- The latter is the expected salary of individuals who all finished college and have since acquired skill level Z=1
- The latter could be expressed as $P(Y_{X=1}|Z_{X=1}=1)$

Example

- U_1, U_2 takes value 0 or 1

Table 4.2 The values attained by $X(u), Y(u), Z(u), Y_0(u), Y_1(u), Z_0(u)$, and $Z_1(u)$ in the model of Eq. (4.7)

$X = u_1 \quad Z = aX + u_2 \quad Y = bZ$								
u_1	u_2	$X(u)$	$Z(u)$	$Y(u)$	$Y_0(u)$	$Y_1(u)$	$Z_0(u)$	$Z_1(u)$
0	0	0	0	0	0	ab	0	a
0	1	0	1	b	b	$(a+1)b$	1	$a+1$
1	0	1	a	ab	0	ab	0	a
1	1	1	$a+1$	$(a+1)b$	b	$(a+1)b$	1	$a+1$

Example

$$E[Y_1|Z=1] = (a+1)*b$$

$$E[Y_0|Z=1] = b$$

$$E[Y|do(X=1), Z=1] = b$$

$$E[Y|do(X=0), Z=1] = b$$

- $E[Y_1 - Y_0|Z=1] = ab \neq 1$

Practical Uses of Counterfactuals

Effect of treatment on the treated

- ETT,

$$ETT = E[Y_1 - Y_0|X = 1]$$

- It can be estimable using observational data, when there exist variables which satisfy backdoor criterion.

Theorem 4.3.1

Theorem 4.3.1

(Counterfactual Interpretation of Backdoor) If a set Z of variables satisfies the backdoor condition relative to (X, Y), then, for all x, the counterfactual Y_x is conditionally independent of X given Z

$$P(Y_x|X, Z) = P(Y_x|Z)$$

- Estimate the probabilities of counterfactuals from observational studies.

$$\begin{aligned} P(Y_x = y) &= \sum_z P(Y_x = y|Z = z)P(z) \\ &= \sum_z P(Y_x = y|Z = z, X = x)P(z) \\ &= \sum_z P(Y = y|Z = z, X = x)P(z) \end{aligned}$$

Recruiment to a Program

- Randomized experiment for figuring out the effect of job training
- Critics says those who self-enroll are more intelligent, more resourceful, ...
- X=1 represent training, Y=1 represent hiring.
- The quauntity that needs to evaluated: ETT,

$$ETT = E[Y_1 - Y_0 | X = 1]$$

- In some situation, (a set Z of covariates which satisfies the backdoor criterion exist), It can be estimated.

Recruiment to a Program

- With modified adjustment formula :

$$\begin{aligned} ETT &= E[Y_1 - Y_0|X = 1] \\ &= E[Y_1|X = 1] - E[Y_0|X = 1] \\ &= E[Y|X = 1] - \sum_z E[Y|X = 0, Z = z]p(Z = z|X = 1) \end{aligned}$$

- Another situation for permitting the identification of ETT occures,
  - Both experimental and nonexperimental data are available, in the form of P(Y=y|do(X=x)) and P(X=x, Y=y)
  - Existence of intermediate variable which satisfies Front-door criterion

Attribution

- 'Probability of necessity'

$$PN = P(Y_0 = 0|X = 1, Y = 1)$$

- 'Probability of sufficency'

$$PS = P(Y_1 = 1|X = 0, Y = 0)$$

- If Y is monotonic relative to X and If P(y|do(x)) is identifiable(randomized or observed with backdoor...).

Thm 4.5.2

Theorem 4.5.2

If Y is monotonic relative to X, that is, $Y_1(u) \geq Y_0(u)$ *for all u, then PN is identifiable whenever the casual effect* $P(y|do(x))$ *is identifiable, and*

$$PN = \frac{P(y) - p(y|do(x^{'})}{P(x,y)}$$

or, substituting $P(y) = P(y|x)P(x) + P(y|x^{'})(1 - P(x))$*, we obtain*

$$PN = \frac{P(y|x) - P(y|x^{'})}{P(y|x)} + \frac{P(y|x) - P(y|do(x^{'}))}{P(x,y)}$$

Personal Decision Making

- Ms Jones receive treatment A and B for tumor, Ten years later, she is alive, and the tumor has not recurred.
- Mrs Smith, on the other hand, receive treatment B alone, her tumor recurred after a year.
- Randomized experiments show receiving treatment A and B is effective than receiving treatment A alone
- However, these were population results. Can we infer from them the specific cases of Ms Jones and Mrs Smith?

Personal Decision Making

- X = 1 represent receive treatment A and B, Y = 1 represent remission of treatment B.
- 'Atrribution'
- For Ms Jones, 'Probability of necessity'

$$PN = P(Y_0 = 0 | X = 1, Y = 1)$$

- For Mrs Smith, 'Probability of sufficency'

$$PS = P(Y_1 = 1 | X = 0, Y = 0)$$

- 'Probalbility of necessity and sufficency'

$$PNS = P(Y_1 = 1, Y_0 = 0)$$

Personal Decision Making

- 'Probalbility of necessity and sufficency'

$$PNS = P(Y_1 = 1, Y_0 = 0)$$

- PNS can be measured if we assume monotonicity.
- Under monotonicity,

$$PNS = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0))$$

- Such quantification of individual risk is extremely important in personal decision making.

END