Journal Club em Causalidade Reunião 3
Causal Diagrams for Empirical Research
Judea Pearl

Widemberg da Silva Nobre, PhD
Pós-doc FAPESP - 2021/06057-0

INSPER
Instituto de Ensino e Pesquisa

08 de outubro de 2021

Introdução

- Neste encontro, o objetivo é formalizar a linguagem causal e trazer à tona questões importantes envolvendo a análise de dados

- Começaremos discutindo estruturas de grafos

- Ao introduzir alguns conceitos introdutórios, esperamos fornecer uma visão geral sobre os cuidados que precisam ser tomados quando desejamos estimar efeitos causais

Conceitos básicos

- Um grafo é um par, $G = (V, E)$, onde V é um conjunto de vértices (variáveis aleatórias) e E é um conjunto de arestas conectando pares de vértices (representando relações probabilísticas)
- Arestas podem ser direcionadas ou não $\Rightarrow$ Se todas as arestas são direcionadas, então o grafo é dito directed graph
- Uma sequência de arestas é chamada path $\Rightarrow$ Dita direcionada se todas suas arestas tem uma flecha saindo do primeiro para o segundo vértice

- Baseado nas paths, um grafo direcionado pode ser classificado como cyclic (se inclui directed cycles) ou acyclic (se não possui directed cycles). Se o grafo é directed e acyclic então ele é chamado directed acyclic graph (DAG)

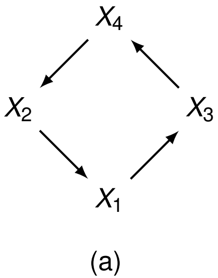

Figure 1: O Painel (a) mostra um exemplo de cyclic graph enquanto o Painel (b) mostra um acyclic grafo.

- Uma causal theory é uma quadrúpla $\mathscr{T} = (V, U, F, \{f_i\})$, onde:

  (i) - $V = \{X_1, \cdots, X_n\}$ é um conjunto de variáveis observadas;

  (ii) - $U = \{U_1, \cdots, U_m\}$ é um conjunto variáveis não-observadas representando ruídos ou distúrbios;

  (iii) - F é uma medida de probabilidade definida sobre U;

  (iv) - $\{f_i\}$ representa um conjunto de n funções determinísticas tal que $X_i = f_i(X_1, \cdots, X_n; U_1, \cdots, U_m)$, $\forall_i$ (Modelo estrutural).

  Essa estrutural funcional possui relação 1-a-1 com a representação gráfica

- Dois pontos de extrema importância na discussão sobre a estimação de efeitos causais

$\Rightarrow$ Explicar as suposições do modelo adotado

$\Rightarrow$ Entender se tais suposições são suficientes para obter estimativas consistentes das quantidades de interesse

Construção do modelo causal

- O primeiro passo da análise é construir um diagrama causal. Pode-se dizer que temos duas formas gerais:

$\Rightarrow$ Construção subjetiva: baseada no conhecimento físico do problema

$\Rightarrow$ Baseada em algorítimos (Causal Discovery)

- Uma vez construído o diagrama, o próximo passo é estabelecer um modelo estrutural referente à estrutura geradora dos dados observados

- O modelo gerador dos dados pode não ser suficiente para estimar, de forma consistente, o efeito causal de interesse. Lembremos do ambiente manipulável que discutimos no nosso último encontro

- Seja $V = \{X_1, \cdots, X_n\}$ uma sequência ordenada de variáveis aleatórias e $f(v)$ uma função densidade de probabilidade tal que

$$f(x_1, \cdots, x_n) = \prod_j f(x_j | x_1, \cdots, x_{j-1}). \tag{1}$$

- Sendo Pa_j o conjunto de variáveis com arestas direcionadas à X_j, temos

$$f(x_j | x_1, \cdots, x_{j-1}) \equiv f(x_j | Pa_j), \tag{2}$$

- Se Pa_j é o menor conjunto de variáveis tal que é possível gerar X_j independente de X_k ($\notin Pa_j$), então Pa_j é chamado de Markovian parents de X_j

- Usando o item (iv) da causal theory, temos

$$X_j = f(Pa_j, u_j), \ (i = 1, \cdots, n) \tag{3}$$

- Essas interpretações nos levam ao conceito de d-separation

Definição (d-separation): Let X, Y and Z be three disjoint subsets of nodes in a directed acyclic graph G, and let p be any path between a node in X and a node in Y, where by "path" we mean any succession of arcs, regardless of their directions. Then Z is said to block p if there is a node w on p satisfying one of the following two conditions:

(i) w has converging arrows along p, and neither w nor any of its descendants are in Z, or,

(ii) w does not have converging arrows along p, and w is in Z

Further, Z is said to d-separate X from Y, in G, written $(X \perp\!\!\!\perp Y|Z)_G$, if and only if Z blocks every path from a node in X to a node in Y

Definição (causal effect): Given two disjoint sets of variables, X and Y, the causal effect of X on Y, denoted $pr(y|\mathrm{do}(X=\tilde{\mathrm{x}}))$, is a function from X to the space of probability distributions on Y For each realisation x of X, $pr(y|\mathrm{do}(X=\tilde{\mathrm{x}}))$ gives the probability of $Y=y$ induced on deleting from the model (3) all equations corresponding to variables in X and substituting x for X in the remainder.

- Se pensarmos numa estrutura gráfica, o efeito causal $pr(y|\tilde{\mathrm{x}}_i)$ sugere remover todos os links entre Pa_i e X_i, enquanto todos os demais links são mantidos
- Esse procedimento é o que chamamos de *do operator* e está associado ao conceito de intervenção
- Note que esse procedimento se aplica a vértices de qualquer natureza física, ou seja, a discussão sobre o que pode ser uma causa não é relevante aqui

Definição (Back-door criterion) . A set of variables Z satisfies the back-door criterion relative to an ordered pair of variables (X_i, X_j) in a directed acyclic graph G if:

(i) no node in Z is a descendant of X_i

(ii) Z blocks every path between X_i and X_j which contains an arrow into X_i. If X and Y are two disjoint sets of nodes in G, Z is said to satisfy the back-door criterion relative to (X, Y) if it satisfies it relative to any pair (X_i, X_j) such that $X_i \in X$ and $X_j \in Y$.

Definição (Identifiability): The causal effect of X on Y is said to be identifiable if the quantity $(X \perp\!\!\!\perp Y|Z)_G$ can be computed uniquely from any positive distribution of the observed variables that is compatible with G.

Exemplificação via estudo simulado

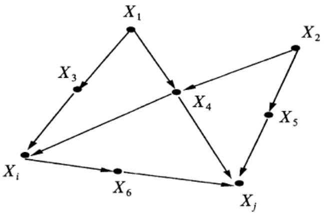

Figure 2: Diagrama ilustrando o critério back-door; ajustando (controlando) para as variáveis $\{X3, X4\}$ ou $\{X4, X5\}$ permite estimar de forma consistente $pr(y|\mathrm{do}(X = \tilde{\mathrm{x}}))$

```
set.seed(12345)
nreps<-1000
nvec<-c(100,200,500)
ests.matM1<-array(0,c(length(nvec),7,nreps))
for(nval in 1:length(nvec)){
 n<-nvec[nval]
  for(irep in 1:nreps){
   X1 <- rnorm(n)
   X2 <- rnorm(n)
   X3 <- X1 + rnorm(n)
   X4 <- X1 + X2 + rnorm(n)
   X5 <- X2 + rnorm(n)
   Xi <- X3 + X4 + rnorm(n)
   X6 <- Xi + rnorm(n)
   Xj <- X4 + X5 + X6 + rnorm(n)
   ests.matM1[nval,1,irep] <- coefficients(lm (Xj ~ Xi))[2]
   ests.matM1[nval,2,irep] <- coefficients(lm (Xj ~ Xi+X4))[2]
   ests.matM1[nval,3,irep] <- coefficients(lm (Xj ~ Xi+X5))[2]
   ests.matM1[nval,4,irep] <- coefficients(lm (Xj ~ Xi+X3))[2]
   ests.matM1[nval,5,irep] <- coefficients(lm (Xj ~ Xi+X4+X5))[2]
   ests.matM1[nval,6,irep] <- coefficients(lm (Xj ~ Xi+X3+X4))[2]
   ests.matM1[nval,7,irep] <- coefficients(lm (Xj ~ Xi+X3+X4+X5))[2]
  }
}
```

	n=100	n=200	n=500
lm (Xj $\sim$ Xi)	1.62	1.62	1.62
lm (Xj $\sim$ Xi+X4)	0.87	0.87	0.87
lm (Xj $\sim$ Xi+X5)	1.47	1.47	1.46
lm (Xj $\sim$ Xi+X3)	2.01	1.99	2.00
lm (Xj $\sim$ Xi+X4+X5)	0.99	1.00	1.00
lm (Xj $\sim$ Xi+X3+X4)	1.00	0.99	1.00
lm (Xj $\sim$ Xi+X3+X4+X5)	1.00	1.00	1.00

Table 1: Resultados do estudo simulado.

- O efeito causal deste exemplo é o average causal effect (ACE), o qual coincide com o coeficiente de regressão usual de X_i para este particular exemplo

- Claramente, o condicionamento somente a X_i leva a um viés na estimativa do ACE (o valor verdadeiro é 1)

- X_1 e X_2 são marginalmente independentes

- O condicionamento a X_4, torna X_1 e X_2 dependentes (X_4 é um collider nessa relação). A path $(X_i, X_3, X_1, X_4, X_2, X_4, X_j)$ é uma back-door path entre X_i e X_j

- O condicionamento em X_3 ou X_5 não controla a back-door path (X_i, X_4, X_j) entre X_i e X_j

- Quando condicionados à $\{X3, X4\}$ or $\{X4, X5\}$ o critério back-door é satisfeito

Exemplo Simulado 1 - Randomized study

- Modelo estrutural de geração dos dados:
  $Z \sim normal(0,1)$; $X \sim normal(0,1)$;
  $Y \sim normal(\tau X + Z, 1)$

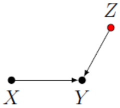

Figure 3: Exemplo 1

- Resultados

	n=100	n=200	n=500
lm (Y $\sim$ X)	5.00	5.00	5.00
lm (Y $\sim$ X + Z)	5.00	5.00	5.00

Table 2: Média de $\hat{\tau}$ sobre 1000 replicações de Monte Carlo. O valor verdadeiro de τ é 5.

Exemplo Simulado 2 - M-bias

- Modelo estrutural de geração dos dados:
  $U_1, U_2 \sim normal(0,1)$, $Z \sim normal(U_1 + U_2, 1)$;
  $X \sim normal(U_1, 1)$; $Y \sim normal(U_2 + \tau X, 1)$

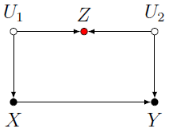

Figure 4: Exemplo 2

- Resultados

	n=100	n=200	n=500
lm (Y $\sim$ X)	5.00	5.00	5.00
lm (Y $\sim$ X + Z)	4.80	4.80	4.80

Table 3: Média de $\hat{\tau}$ sobre 1000 replicações de Monte Carlo. O valor verdadeiro de τ é 5.

Exemplo Simulado 3 - Bias Amplification

- Modelo estrutural de geração dos dados:
  $, U, Z \sim normal(0,1); X \sim normal(U + Z, 1)$;
  $Y \sim normal(5X + U, 1)$

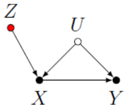

Figure 5: Exemplo 3

- Resultados

	n=100	n=200	n=500
lm (Y $\sim$ X)	5.34	5.34	5.33
lm (Y $\sim$ X + Z)	5.50	5.50	5.50

Table 4: Média de $\hat{\tau}$ sobre 1000 replicações de Monte Carlo. O valor verdadeiro de τ é 5.

Exemplo Simulado 4 - Selection Bias

- Modelo estrutural de geração dos dados:
  $U, X \sim normal(0,1)$; $Z \sim normal(U + X, 1)$;
  $Y \sim normal(5X + U, 1)$

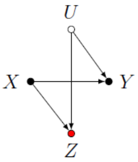

Figure 6: Exemplo 4

- Resultados

	n=100	n=200	n=500
lm (Y ∼ X)	5.01	5.00	5.00
lm (Y ∼ X + Z)	4.51	4.50	4.50

Table 5: Média de $\hat{\tau}$ sobre 1000 replicações de Monte Carlo. O valor verdadeiro de τ é 5.

Resumo

- Nós vimos um pouco da formalização do paradigma causal em termos de DAGs e modelos estruturais
- Vimos que é necessário controlar algumas covariáveis que induzem confundimento (critério back-door)
- Vimos também que é a inclusão de algumas covariáveis pode induzir viés
- Conclusão: sem uma discussão minunciosa dos dados e da plausibilidade de determinadas suposições, conclusões a respeito de parâmetros causais podem estar viesadas

- No próximo encontro iremos falar sobre o propensity score, que é uma ferramenta bastante útil na análise de dados observacionais
- Nos exemplos de hoje, trabalhamos estruturas bem simples, com poucas covariáveis, e com uma simples determinação do modelo resposta. No próximo encontro falaremos de cenários com estruturas complexas, onde o propensity score pode ser mais útil
- Doubly robust methods