

Inferência Causal

Notas de Aula

Rafael Bassi Stern

Última revisão: 27 de agosto de 2024

Por favor, enviem comentários, typos e erros para rbstern@gmail.com

Agradecimentos: Vitor Mello

“Teaching is giving opportunities to students to discover things by themselves.”

George Pólya

Conteúdo

1. Por que estudar Inferência Causal?	7
1.1. O Paradoxo de Simpson	7
1.1.1. Exercícios	8
2. Modelo Causal (CM)	11
2.1. Elementos de Modelos Probabilísticos em Grafos	11
2.1.1. Grafo Direcionado	11
2.1.2. Grafo Direcionado Acíclico (DAG)	12
2.1.3. Modelo Probabilístico em um DAG	13
2.1.4. Exemplos de Modelo Probabilístico em um DAG	15
Confundidor (Confounder)	15
Cadeia (Chain)	15
Colisor (Collider)	16
2.1.5. Modelo Causal (Causal Model)	18
2.1.6. Exercícios	19
2.2. Independência Condicional e D-separação	20
2.2.1. Independência Condicional	20
2.2.2. D-separação	20
2.2.3. Exercícios	23
3. Intervenções	25
3.1. O modelo de probabilidade para intervenções	25
3.1.1. Exercícios	30
3.2. Controlando confundidores (critério <i>backdoor</i>)	31
3.2.1. Identificação causal usando o critério <i>backdoor</i>	33
3.2.2. Estimação usando o critério <i>backdoor</i>	34
Fórmula do ajuste	34
Ponderação pelo inverso do escore de propensidade (IPW)	38
Estimador duplamente robusto	40
3.2.3. Exercícios	41
3.2.4. Regression Discontinuity Design (RDD)	42
Identificação causal no RDD	42
Estimação no RDD	43
3.2.5. Exercícios	45
3.3. Controlando mediadores (critério <i>frontdoor</i>)	45
Identificação causal	46
Estimação pelo critério <i>frontdoor</i>	47
	5

3.4. Do-calculus	47
3.4.1. Exercícios	48
4. Resultados potenciais	49
4.1. Levando a intuição do SCM ao POM	51
4.1.1. Exercícios	53
4.2. Variáveis Instrumentais	53
4.3. Contrafactuais	55
4.3.1. Contrafactuais e Responsabilidade Civil	56
4.3.2. Exercícios	58
5. Descoberta Causal	59
5.1. Identificabilidade na Descoberta Causal	59
5.2. Algoritmos de Descoberta Causal	60
5.2.1. Algoritmos baseados em testes de independência condicional	60
Algoritmo SGS	60
Algoritmo PC	60
A. Demonstrações	63
A.1. Seção 2.1.4 (Exemplos de Modelo Probabilístico em um DAG)	63
A.2. Seção 2.1.5 (Modelo Causal (Causal Model))	64
A.3. Seção 2.2 (Independência Condicional e D-separação)	65
A.3.1. Relativas ao Lema 2.45	65
A.3.2. Teorema 2.49	67
A.4. Relativas à Seção 3.1 (O modelo de probabilidade para intervenções)	68
A.4.1. Relativas ao Teorema 3.6	68
A.5. Seção 3.2 (Controlando confundidores (critério <i>backdoor</i>))	69
A.5.1. Teorema 3.19	69
A.5.2. Teoremas 3.21 e 3.22	72
A.5.3. Teorema 3.26	72
A.5.4. Teorema 3.29	73
A.5.5. Teorema 3.32	74
A.5.6. Teorema 3.39	75
A.6. Seções 3.3 e 3.4 (Controlando mediadores (critério <i>frontdoor</i>))	76
A.6.1. Teorema 3.48	76
A.6.2. Teorema 3.44	77
A.6.3. Teorema 3.45	78
A.7. Seção 4.1 (Levando a intuição do SCM ao POM)	78
A.8. Seção 4.2 (Variáveis Instrumentais)	80
A.9. Seção 4.3 (Contrafactuais)	83
A.10. Seção 5.1 (Identificabilidade na Descoberta Causal)	83

1. Por que estudar Inferência Causal?

Você já deve ter ouvido diversas vezes que **correlação não implica causalidade**. Contudo, o que é causalidade e como ela pode ser usada para resolver problemas práticos? Antes de estudarmos definições formais, veremos como conceitos intuitivos de causalidade podem ser necessários para resolver questões usuais em Inferência Estatística. Para tal, a seguir estudaremos um exemplo de ?.

1.1. O Paradoxo de Simpson

Considere que observamos em 700 pacientes 3 variáveis: T e C são as indicadoras de que, respectivamente, o paciente recebeu um tratamento e o paciente curou de uma doença, e Z é uma variável binária cujo significado será discutido mais tarde. Os dados foram resumidos na [tabela 1.1](#).

Em uma primeira análise desta tabela, podemos verificar a efetividade do tratamento dentro de cada valor de Z . Por exemplo, quando $Z = 0$, a frequência de recuperação dentre aqueles que receberam e não receberam o tratamento são, respectivamente: $\frac{81}{6+81} \approx 0.93$ e $\frac{234}{36+234} \approx 0.87$. Similarmente, quando $Z = 1$, as respectivas frequências são: $\frac{192}{71+192} \approx 0.73$ e $\frac{55}{25+55} \approx 0.69$. À primeira vista, para todos os valores de Z , a taxa de recuperação é maior com o tratamento do que sem ele. Isso nos traz informação de que o tratamento é efetivo na recuperação do paciente?

Em uma segunda análise, podemos considerar apenas as contagens para as variáveis T e C , sem estratificar por Z . Dentre os pacientes que receberam e não receberam o tratamento as taxas de recuperação são, respectivamente: $\frac{81+192}{6+71+81+192} \approx 0.78$ e $\frac{234+55}{36+25+234+55} \approx 0.83$. Isto é, sem estratificar por Z , a frequência de recuperação é maior dentre aqueles que não receberam o tratamento do que dentre aqueles que o receberam.

O que é possível concluir destas análises? Uma conclusão ingênua poderia ser a de que, se Z não for observada, então o tratamento não é recomendado. Por outro lado, se Z é observada, não importa qual seja o seu valor, o tratamento será recomendado. A falta de sentido desta conclusão ingênua é o que tornou este tipo de dado famoso como sendo um caso de Paradoxo de Simpson (?).

Contudo, se a conclusão ingênua é paradoxal e incorreta, então qual conclusão pode ser obtida destes dados? A primeira lição que verificaremos é que não é possível obter uma conclusão sobre o **efeito causal** do tratamento usando apenas a informação na tabela, isto é, associações. Para tal, analisaremos a tabela dando dois nomes distintos para a variável Z . Veremos que, usando exatamente os mesmos dados, uma conclusão válida diferente

##	C	0	1
##	Z	T	
##	0	0	36 234
##	1	6	81
##	1	0	25 55
##	1	71	192

Tabela 1.1.: Frequência conjunta das variáveis binárias T , C , e Z .

é obtida para cada nome de Z . Em outras palavras, o efeito causal depende de mais informação do que somente aquela disponível na tabela.

Em um primeiro cenário, considere que Z é a indicadora de que o sexo do paciente é masculino. Observando a tabela, notamos que, proporcionalmente, mais homens receberam o tratamento do que mulheres. Como o tratamento não tem qualquer influência sobre o sexo do paciente, podemos imaginar um cenário em que, proporcionalmente, mais homens escolheram receber o tratamento do que mulheres.

Usando esta observação, podemos fazer sentido do Paradoxo anteriormente obtido. Quando agregamos os dados, notamos que o primeiro grupo de pacientes que receberam o tratamento é predominantemente composto por homens e, similarmente, o segundo grupo de pacientes que não receberam o tratamento é predominantemente composto por mulheres. Isto é, na análise dos dados agregados estamos essencialmente comparando a taxa de recuperação de homens que receberam o tratamento com a de mulheres que não receberam o tratamento. Se assumirmos que, independentemente do tratamento, mulheres tem uma probabilidade de recuperação maior do que homens, então a taxa de recuperação menor no primeiro grupo pode ser explicada pelo fato de ele ser composto predominantemente por homens e não pelo fato de ser o grupo de pacientes que recebeu o tratamento. Também, da análise anterior, obtemos que para cada sexo, a taxa de recuperação é maior com o tratamento do que sem ele. Isto é, neste cenário, o tratamento parece efetivo para a recuperação dos pacientes. Isto significa que a análise estratificando Z é sempre a correta?

Caso o significado da variável Z seja outro, veremos que esta conclusão é incorreta. Considere que Z é a indicadora de que a pressão sanguínea do paciente está elevada. Além disso, é sabido que o tratamento tem como efeito colateral aumentar o risco de pressão elevada nos pacientes. Neste caso, o fato de que há mais indivíduos com pressão elevada dentre aqueles que receberam o tratamento é um efeito direto do tratamento.

Usando esta observação, podemos chegar a outras conclusões sobre o efeito do tratamento sobre a recuperação dos pacientes. Para tal, considere que o tratamento tem um efeito positivo moderado sobre a recuperação dos pacientes, mas que a pressão sanguínea elevada prejudica gravemente a recuperação. Quando fazemos comparações apenas dentre indivíduos com pressão alta ou apenas dentre indivíduos sem pressão alta, não é possível identificar o efeito colateral do tratamento. Isto é, observamos apenas o efeito positivo moderado que o tratamento tem sobre a recuperação. Por outro lado, quando fazemos a análise agregada, observamos que a frequência de recuperação é maior dentre os indivíduos que não receberam o tratamento do que dentre os que o receberam. Isso ocorre pois o efeito colateral negativo tem um impacto maior sobre a recuperação do paciente do que o efeito geral benéfico. Assim, neste cenário, o tratamento não é eficiente para levar à recuperação do paciente.

Como nossas conclusões dependem de qual história adotamos, podemos ver que a mera apresentação da tabela é insuficiente para determinar a eficiência do tratamento. Observando com cuidado os cenários, identificamos uma explicação geral para as diferentes conclusões. No primeiro cenário, quando Z é sexo, Z é uma causa do indivíduo receber ou não o tratamento. Já no segundo cenário, quando Z é pressão elevada, o tratamento é causa de Z . Isto é, a diferença nas relações entre as variáveis explica as diferenças entre as conclusões obtidas.

Ao longo do curso, desenvolveremos ferramentas para formalizar a diferença entre estes cenários e, com base nisso, conseguir estimar o efeito causal que uma variável X tem sobre outra variável Y . Contudo, para tal, será necessário desenvolver um modelo em que seja possível descrever relações causais. Esta questão será tratada no [capítulo 2](#).

1.1.1. Exercícios

Exercício 1.1 (?[p.6]). Há evidência de que há correlação positiva entre uma pessoa estar atrasada e estar apressada. Isso significa que uma pessoa pode evitar atrasos se não tiver pressa? Justifique sua resposta em 8

palavras.

2. Modelo Causal (CM)

No [capítulo 1](#) vimos que as relações causais entre variáveis são essenciais para conseguirmos determinar o efeito que uma variável pode ter em outra. Contudo, como podemos especificar relações causais formalmente?

Como resposta a esta pergunta iremos definir o Modelo Causal (CM), que permite especificar formalmente relações causais. Para tal, será necessário primeiro introduzir modelos probabilísticos em grafos. Um curso completo sobre estes modelos pode ser encontrado, por exemplo, em [?](#). A seguir, estudaremos resultados essenciais destes modelos.

2.1. Elementos de Modelos Probabilísticos em Grafos

2.1.1. Grafo Direcionado

Definição 2.1. Um **grafo direcionado**, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, é composto por um conjunto de vértices, $\mathcal{V} = \{V_1, \dots, V_n\}$, e um conjunto de arestas, $\mathcal{E} = \{E_1, \dots, E_m\}$, onde cada aresta é um par ordenado de vértices, isto é, $E_i \in \mathcal{V}^2$.

Para auxiliar nossa intuição sobre a [Definição 2.1](#), é comum representarmos o grafo por meio de uma figura. Nesta, representamos cada vértice por meio de um ponto. Além disso, para cada aresta, (V_i, V_j) , traçamos uma seta que aponta de V_i para V_j .

Por exemplo, considere que os vértices são $\mathcal{V} = \{V_1, V_2, V_3\}$ e as arestas são $\mathcal{E} = \{(V_1, V_2), (V_1, V_3), (V_2, V_3)\}$. Neste caso, teremos os 3 pontos como vértices e, além disso, traçaremos setas de V_1 para V_2 e para V_3 e, também, de V_2 para V_3 . Podemos desenhar este grafo utilizando os pacotes *dagitty* e *ggdag* (??):

```
library(dagitty)
library(ggdag)
library(ggplot2)

# Especificar o grafo
grafo <- dagitty("dag {
  V1 -> { V2 V3 }
  V2 -> V3
}")

# Exibir a figura do grafo
ggdag(grafo, layout = "circle") +
  theme(axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank()) +
  xlab("") + ylab("")
```

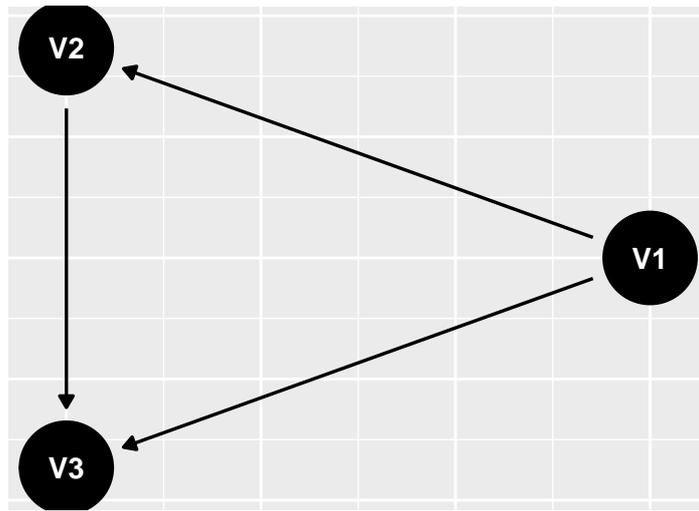


Figura 2.1.: Exemplo de grafo.

Grafos direcionados serão úteis para representar causalidade pois seus vértices serão variáveis e suas arestas irão apontar de cada causa imediata para seu efeito. Por exemplo, no [Capítulo 1](#) consideramos um caso em que Sexo e Tratamento são causas imediatas de recuperação e, além disso, Sexo é causa imediata de Tratamento. O grafo na [figur 2.1](#) poderia representar estas relações se definirmos que V_1 é Sexo, V_2 é Tratamento e V_3 é Recuperação.

Usando a representação de um grafo, podemos imaginar caminhos sobre ele. Um **caminho direcionado** inicia-se em um determinado vértice e, seguindo a direção das setas, vai de um vértice para outro. Por exemplo, (V_1, V_2, V_3) é um caminho direcionado na [figur 2.1](#), pois existe uma seta de V_1 para V_2 e de V_2 para V_3 . É comum denotarmos este caminho direcionado por $V_1 \rightarrow V_2 \rightarrow V_3$. Similarmente, (V_1, V_3, V_2) não é um caminho direcionado, pois não existe seta de V_3 para V_2 . A definição de caminho direcionado é formalizada a seguir:

Definição 2.2. Um **caminho direcionado** é uma sequência de vértices em um grafo direcionado, $C = \{C_1, \dots, C_n\}$ tal que, para cada $1 \leq i < n$, $(C_i, C_{i+1}) \in \mathcal{E}$.

Definição 2.3. Dizemos que V_2 é descendente de V_1 se existe um caminho direcionado de V_1 em V_2 .

Um *caminho* é uma generalização de caminho direcionado. Em um caminho, começamos em um vértice e, seguindo por setas, mas não necessariamente na direção em que elas apontam, vamos de um vértice para outro. Por exemplo, na [figur 2.1](#) vimos que (V_1, V_3, V_2) não é um caminho direcionado pois não existe seta de V_3 para V_2 . Contudo, (V_1, V_3, V_2) é um caminho pois existe uma seta ligando V_3 e V_2 , a seta que aponta de V_2 para V_3 . É comum representarmos este caminho por $V_1 \rightarrow V_3 \leftarrow V_2$. Caminho é formalizado a seguir:

Definição 2.4. Dizemos que vértices V_1 e V_2 são **adjacentes** se $(V_1, V_2) \in \mathcal{E}$ ou $(V_2, V_1) \in \mathcal{E}$.

Definição 2.5. Um **caminho** é uma sequência de vértices, $C = \{C_1, \dots, C_n\}$ tal que, para cada $1 \leq i < n$, C_i e C_{i+1} são adjacentes.

2.1.2. Grafo Direcionado Acíclico (DAG)

Um DAG é um grafo direcionado tal que, para todo vértice, V , não é possível seguir setas partindo de V e voltar para V . Este conceito é formalizado a seguir:

Definição 2.6. Um **grafo direcionado acíclico** (DAG) é um grafo direcionado, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, tal que, para todo vértice, $V \in \mathcal{V}$, não existe um caminho direcionado, $C = \{C_1, \dots, C_n\}$ tal que $C_1 = V = C_n$.

Usualmente representaremos as relações causais por meio de um DAG. Especificamente, existirá uma aresta de V_1 para V_2 para indicar que V_1 é causa imediata de V_2 . Caso um grafo direcionado não seja um DAG, então existe um caminho de V em V , isto é, V seria uma causa de si mesma, o que desejamos evitar.

Um DAG induz uma *ordem parcial* entre os seus vértices. Isto é, se existe uma aresta de V_1 para V_2 , então podemos interpretar que V_1 antecede V_2 causalmente. Com base nesta ordem parcial, é possível construir diversas definições que nos serão úteis.

Dizemos que V_1 é pai de V_2 em um DAG, \mathcal{G} , se existe uma aresta de V_1 a V_2 , isto é, $(V_1, V_2) \in \mathcal{E}$. Denotamos por $Pa(V)$ o conjunto de todos os pais de V . Similarmente $Pa(\mathbb{V})$ é o conjunto de vértices que são pais de algum vértice em \mathbb{V} :

Definição 2.7. O conjunto de **pais** de $\mathbb{V} \subseteq \mathcal{V}$ em um DAG, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, é:

$$Pa(\mathbb{V}) := \{V^* \in \mathcal{V} : \exists V \in \mathbb{V} \text{ tal que } (V^*, V) \in \mathcal{E}\}.$$

Similarmente, dizemos que V_1 é um ancestral de V_2 em um DAG, se V_1 antecede V_2 causalmente. Isto é, se V_1 é pai de V_2 ou, pai de pai de V_2 , ou pai de pai de pai de V_2 , e assim por diante ... Denotamos por $Anc(\mathbb{V})$ o conjunto de todos os ancestrais de elementos de \mathbb{V} :

Definição 2.8. Em um DAG, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, o conjunto de **ancestrais** de $\mathbb{V} \subseteq \mathcal{V}$, $Anc(\mathbb{V})$, é tal que $Anc(\mathbb{V}) \subseteq \mathcal{V}$ e $V^* \in Anc(\mathbb{V})$ se e somente se existe $V \in \mathbb{V}$ e um caminho direcionado, C , tal que $C_1 = V^*$ e $C_i = V$.

Note que podemos interpretar $Anc(\mathbb{V})$ como o conjunto de todas as causas diretas e indiretas de \mathbb{V} .

Finalmente, diremos que um conjunto de vértices, $\mathcal{A} \subseteq \mathcal{V}$ é *ancestral* em um DAG, se não existe algum vértice fora de \mathcal{A} que seja pai de algum vértice em \mathcal{A} . Segundo nossa interpretação causal, \mathcal{A} será ancestral quando nenhum vértice fora de \mathcal{A} é causa direta de algum vértice em \mathcal{A} :

Definição 2.9. Dizemos que $\mathcal{A} \subseteq \mathcal{V}$ é **ancestral** em um DAG se, para todo vértice $V \in \mathcal{A}$, temos que $Pa(V) \subseteq \mathcal{A}$.

Lema 2.10. Em um DAG, \mathcal{G} , para todo $\mathbb{V} \subseteq \mathcal{V}$, $Anc(\mathbb{V})$ é ancestral.

2.1.3. Modelo Probabilístico em um DAG

Um modelo probabilístico em um DAG é tal que cada um dos vértices é uma variável aleatória. O DAG será usado para descrever relações de independência condicional existentes entre estas variáveis. Mais especificamente, cada vértice será independente dos demais vértices dados os seus pais. Uma maneira alternativa de pensar sobre esta afirmação é imaginar que cada vértice é gerado somente pelos seus pais. Esta intuição é formalizada em [Definição 2.11](#):

Definição 2.11. Para \mathcal{V} um conjunto de variáveis aleatórias, dizemos que uma função de densidade sobre \mathcal{V} , f , é compatível com um DAG, \mathcal{G} , se:

$$f(v_1, \dots, v_n) = \prod_{i=1}^n f(v_i | Pa(v_i)).$$

Quando não há ambiguidade, também dizemos que \mathcal{G} é compatível com f neste caso.

Exemplo 2.12. Considere que $X \sim \text{Bernoulli}(0.5)$, $Y|X = 1 \sim \text{Bernoulli}(0.99)$ e $Y|X = 0 \sim \text{Bernoulli}(0.01)$. Neste caso,

$$\begin{aligned} f(Y = 1) &= f(X = 0, Y = 1) + f(X = 1, Y = 1) \\ &= f(X = 0)f(Y = 1|X = 0) + f(X = 1)f(Y = 1|X = 1) \\ &= 0.5 \cdot 0.01 + 0.5 \cdot 0.99 = 0.5 \end{aligned}$$

Como $f(X = 1, Y = 1) = 0.5 \cdot 0.99 \neq 0.5 \cdot 0.5 = f(X = 1)f(Y = 1)$, decorre da [Definição 2.11](#) que f não é compatível com o DAG sem arestas em que $\mathcal{V} = \{X, Y\}$. Em outras palavras, X e Y não são independentes. Como sempre é verdade que $f(x, y) = f(x)f(y|x)$ e que $f(x, y) = f(y)f(x|y)$, f é compatível com os DAGs $X \rightarrow Y$ e com $X \leftarrow Y$.

Exemplo 2.13. Considere que $f(x, y) = f(x)f(y)$. Isto é, (X, Y) são independentes segundo f . Neste caso, f é compatível com qualquer DAG sobre $\mathcal{V} = \{X, Y\}$.

Quando \mathcal{V} tem muitos elementos, pode ser difícil verificar se a [Definição 2.11](#) está satisfeita. Para esses casos, pode ser útil aplicar o [Lema 2.14](#):

Lema 2.14. Uma função de densidade, f , é compatível com um DAG, \mathcal{G} , se e somente se, existem funções, g_1, \dots, g_n tais que:

$$f(v_1, \dots, v_n) = \prod_{i=1}^n g_i(v_i, Pa(v_i)), \quad e \quad \int g_i(v_i, Pa(v_i)) dv_i = 1$$

Exemplo 2.15. Considere que

$$f(x_1, x_2, x_3) = 0.5 \cdot 0.9^{\mathbb{I}(x_1=x_2)} \cdot 0.1^{\mathbb{I}(x_1 \neq x_2)} \cdot 0.8^{\mathbb{I}(x_2=x_3)} \cdot 0.2^{\mathbb{I}(x_2 \neq x_3)}.$$

Tome $\mathcal{G} = X_1 \rightarrow X_2 \rightarrow X_3$. Para \mathcal{G} , $Pa(X_1) = \emptyset$, $Pa(X_2) = \{X_1\}$ e $Pa(X_3) = \{X_2\}$. Assim, tomando $g_1(x_1, Pa(x_1)) = 0.5$, $g_2(x_2, Pa(x_2)) = 0.9^{\mathbb{I}(x_1=x_2)} \cdot 0.1^{\mathbb{I}(x_1 \neq x_2)}$ e $g_3(x_3, Pa(x_3)) = 0.8^{\mathbb{I}(x_2=x_3)} \cdot 0.2^{\mathbb{I}(x_2 \neq x_3)}$, temos que

$$f(x_1, x_2, x_3) = g_1(x_1, Pa(x_1)) \cdot g_2(x_2, Pa(x_2)) \cdot g_3(x_3, Pa(x_3))$$

Isto é, decorre do [Lema 2.14](#) que f é compatível com \mathcal{G} .

Exercício 2.16. Usando a mesma f do [Exemplo 2.15](#), prove que f é compatível com o DAG $X_1 \leftarrow X_2 \leftarrow X_3$. Temos que f é compatível com quais outros DAG's?

Se \mathcal{A} é ancestral em um DAG, então $f(\mathcal{A})$ pode ser decomposto de forma similar a $f(\mathcal{V})$. Este fato será útil e é formalizado no [Lema 2.17](#).

Lema 2.17. Seja $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ um DAG. Se \mathcal{A} é ancestral e f é compatível com \mathcal{G} , então

$$f(\mathcal{A}) = \prod_{V \in \mathcal{A}} f(V|Pa(V))$$

A seguir, estudaremos três tipos fundamentais de modelos probabilísticos em DAG's com 3 vértices. A intuição obtida a partir destes exemplos continuará valendo quando estudarmos grafos mais gerais.

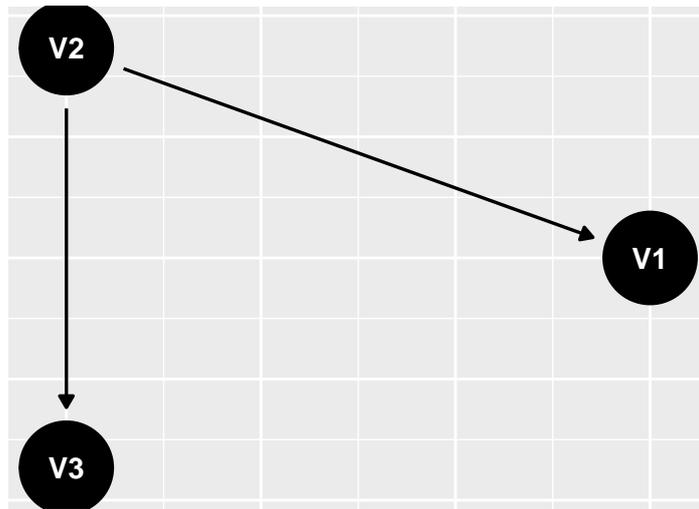


Figura 2.2.: Ilustração de confundidor.

2.1.4. Exemplos de Modelo Probabilístico em um DAG

Nos exemplos a seguir, considere que $\mathcal{V} = (V_1, V_2, V_3)$.

Confundidor (Confounder)

No modelo de confundidor, as únicas duas arestas são (V_2, V_1) e (V_2, V_3) . Uma ilustração de um confundidor pode ser encontrada na [figur 2.2](#). O modelo de confundidor pode ser usado quando acreditamos que V_2 é uma causa comum a V_1 e a V_3 . Além disso, V_1 não é causa imediata de V_3 nem vice-versa.

Em um modelo de confundidor a relação de dependência entre V_1 e V_3 é explicada pelos resultados a seguir:

Lema 2.18. *Para qualquer probabilidade compatível com o DAG na [figur 2.2](#), $V_1 \perp V_3 | V_2$.*

Lema 2.19. *Existe ao menos uma probabilidade compatível com o DAG na [figur 2.2](#) tal que $V_1 \not\perp V_3$.*

Combinando os [Lemas 2.18](#) e [2.19](#) é possível compreender melhor como usaremos confundidores num contexto causal. Nestes casos, V_2 será uma causa comum a V_1 e a V_3 . Esta causa comum torna V_1 e V_3 associados, ainda que nenhum seja causa direta ou indireta do outro.

Podemos contextualizar estas ideias em um caso de diagnóstico de dengue. Considere que V_2 é a indicadora de que um indivíduo tem dengue, e V_1 e V_3 são indicadoras de sintomas típicos de dengue, como dor atrás dos olhos e febre. Neste caso, V_1 e V_3 tipicamente são associados: caso um paciente tenha febre, aumenta a probabilidade de que tenha dengue e, portanto, aumenta a probabilidade de que tenha dor atrás dos olhos. Contudo, apesar dessa associação V_3 não tem influência causal sobre V_1 . Se aumentarmos a temperatura corporal do indivíduo, não aumentará a probabilidade de que ele tenha dor atrás dos olhos. A dengue que causa febre, não o contrário.

Cadeia (Chain)

No modelo de cadeia, as únicas duas arestas são (V_1, V_2) e (V_2, V_3) . Uma ilustração de uma cadeia pode ser encontrada na [figur 2.3](#). Neste modelo, acreditamos que V_1 é causa de V_2 que, por sua vez, é causa de V_3 . Assim, V_1 é ancestral de V_3 , isto é, o primeiro é causa indireta do segundo.

Em um modelo de cadeia a relação de dependência entre V_1 e V_3 é explicada pelos resultados a seguir:

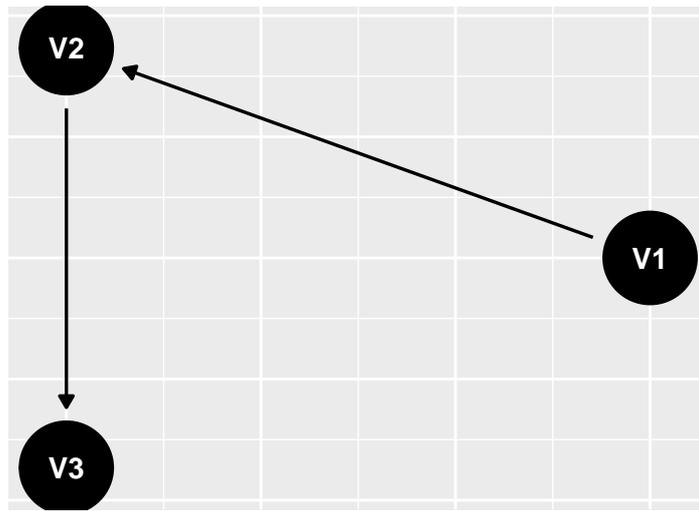


Figura 2.3.: Ilustração de cadeia.

Lema 2.20. Para qualquer probabilidade compatível com o DAG na [figur 2.3](#), $V_1 \perp\!\!\!\perp V_3|V_2$.

Lema 2.21. Existe ao menos uma probabilidade compatível com o DAG na [figur 2.3](#) tal que $V_1 \not\perp\!\!\!\perp V_3$.

Combinando os [Lemas 2.20](#) e [2.21](#) é possível compreender melhor como usaremos cadeias num contexto causal. Nestes casos, V_2 será uma consequência de V_1 e uma causa de V_3 . Assim, a cadeia torna V_1 e V_3 e associados, ainda que nenhum seja causa direta do outro. Contudo, ao contrário do confundidor, neste caso V_1 é uma causa indireta de V_3 , isto é, tem influência causal sobre V_3 .

Para contextualizar estas ideias, considere que V_1 é a indicadora de consumo elevado de sal, V_2 é a indicadora de pressão alta, e V_3 é a indicadora de ocorrência de um derrame. Como consumo elevado de sal causa pressão alta e pressão alta tem influência causal sobre a ocorrência de um derrame, pressão alta é uma cadeia que é um mediador entre consumo elevado de sal e ocorrência de derrame. Assim, consumo elevado de sal tem influência causal sobre a ocorrência de derrame.

Colisor (Collider)

O último exemplo de DAG com 3 vértices que estudaremos é o de modelo de colisor, em que as únicas duas arestas são (V_1, V_2) e (V_3, V_2) . Uma ilustração de um colisor pode ser encontrada na [figur 2.4](#). O modelo de colisor pode ser usado quando acreditamos que V_1 e V_3 são causas comuns a V_2 . Além disso, V_1 não é causa imediata de V_3 nem vice-versa.

Em um modelo de colisor a relação de dependência entre V_1 e V_3 é explicada pelos resultados a seguir:

Lema 2.22. Para qualquer probabilidade compatível com o DAG na [figur 2.4](#), $V_1 \perp\!\!\!\perp V_3$.

Lema 2.23. Existe ao menos uma probabilidade compatível com o DAG na [figur 2.4](#) tal que $V_1 \not\perp\!\!\!\perp V_3|V_2$.

Combinando os [Lemas 2.22](#) e [2.23](#) vemos como utilizaremos confundidores num contexto causal. Nestes casos, V_1 e V_3 serão causas comuns e independentes de V_2 . Uma vez que obtemos informação sobre o efeito comum, V_2 , V_1 e V_3 passam a ser associados.

Esse modelo pode ser contextualizado observando a prevalência de doenças em uma determinada população (?). Considere que V_1 e V_3 são indicadoras de que um indivíduo tem doenças que ocorrem independentemente na

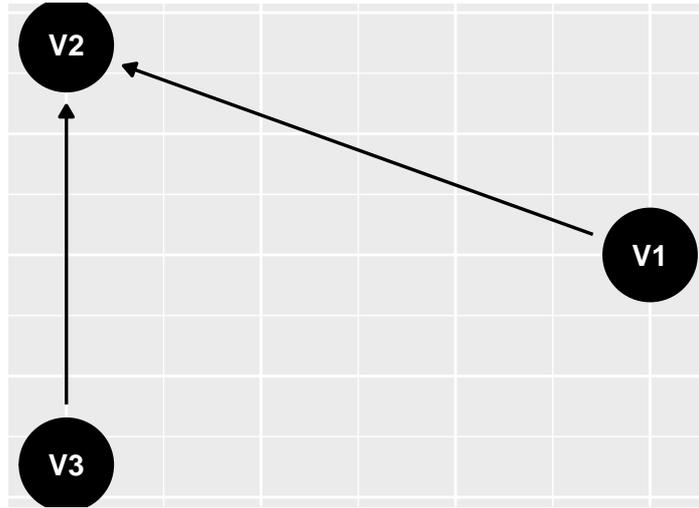


Figura 2.4.: Ilustração de colisor.

população. Além disso, V_2 é a indicadora de que o indivíduo foi hospitalizado, isto é, V_2 é influenciado causalmente tanto por V_1 quanto por V_3 . Para facilitar as contas envolvidas, desenvolveremos o exemplo com distribuições fictícias. Considere que V_1 e V_3 são independentes e tem distribuição Bernoulli(0.05). Além disso, quanto maior o número de doenças, maior a probabilidade de o indivíduo ser hospitalizado. Por exemplo, $\mathbb{P}(V_2 = 1|V_1 = 0, V_3 = 0) = 0.01$, $\mathbb{P}(V_2 = 1|V_1 = 0, V_3 = 1) = 0.1$, $\mathbb{P}(V_2 = 1|V_1 = 1, V_3 = 0) = 0.1$, e $\mathbb{P}(V_2 = 1|V_1 = 1, V_3 = 1) = 0.5$.

Com base nestas especificações, podemos verificar se V_1 e V_3 estão associados quando $V_2 = 1$. Para tal, primeiramente calcularemos algumas probabilidades conjuntas que serão úteis:

$$\begin{cases} \mathbb{P}(V_1 = 0, V_2 = 1, V_3 = 0) &= 0.95 \cdot 0.01 \cdot 0.95 = 0.009025 \\ \mathbb{P}(V_1 = 0, V_2 = 1, V_3 = 1) &= 0.95 \cdot 0.1 \cdot 0.05 = 0.0475 \\ \mathbb{P}(V_1 = 1, V_2 = 1, V_3 = 0) &= 0.05 \cdot 0.1 \cdot 0.95 = 0.0475 \\ \mathbb{P}(V_1 = 1, V_2 = 1, V_3 = 1) &= 0.05 \cdot 0.5 \cdot 0.05 = 0.00125 \end{cases} \quad (2.1)$$

Com base nestes cálculos é possível obter a prevalência da doença dentre os indivíduos hospitalizados:

$$\begin{aligned} \mathbb{P}(V_1 = 1|V_2 = 1) &= \frac{\mathbb{P}(V_1 = 1, V_2 = 1)}{\mathbb{P}(V_2 = 1)} \\ &= \frac{0.0475 + 0.00125}{0.009025 + 0.0475 + 0.0475 + 0.00125} && \text{linking (2.1)} \\ &\approx 0.46 \end{aligned}$$

Finalmente,

$$\begin{aligned}
\mathbb{P}(V_1 = 1|V_2 = 1, V_3 = 1) &= \frac{\mathbb{P}(V_1 = 1, V_2 = 1, V_3 = 1)}{\mathbb{P}(V_2 = 1, V_3 = 1)} \\
&= \frac{\mathbb{P}(V_1 = 1, V_2 = 1, V_3 = 1)}{\mathbb{P}(V_1 = 0, V_2 = 1, V_3 = 1) + \mathbb{P}(V_1 = 1, V_2 = 1, V_3 = 1)} \\
&= \frac{0.00125}{0.0475 + 0.00125} && \text{likning (2.1)} \\
&\approx 0.26
\end{aligned}$$

Como $\mathbb{P}(V_1 = 1|V_2 = 1) = 0.46 \neq 0.26 \approx \mathbb{P}(V_1 = 1|V_2 = 1, V_3 = 1)$, verificamos que V_1 não é independente de V_3 dado V_2 . De fato, ao observar que um indivíduo está hospitalizado e tem uma das doenças, a probabilidade de que ele tenha a outra doença é inferior àquela obtida se soubéssemos apenas que o indivíduo está hospitalizado.

Esta observação não implica que uma doença tenha influência causal sobre a outra. Note que a frequência de hospitalização aumenta drasticamente quando um indivíduo tem ao menos uma das doenças. Além disso, cada uma das doenças é relativamente rara na população geral. Assim, dentre os indivíduos hospitalizados, a frequência daqueles que tem somente uma das doenças é maior do que seria caso as doenças não estivessem associadas. Quando fixamos o valor de uma consequência comum (hospitalização), as causas (doenças) passam a ser associadas. Esta associação não significa que infectar um indivíduo com uma das doenças reduz a probabilidade que ele tenha a outra.

2.1.5. Modelo Causal (Causal Model)

Com base nos conceitos abordados anteriormente, finalmente podemos definir o Modelo Causal (CM):

Definição 2.24. Um CM é um par (\mathcal{G}, f) tal que $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ é um DAG (Definição 2.6) e f é uma função de densidade sobre \mathcal{V} compatível com \mathcal{G} (Definição 2.11). Neste caso, é comum chamarmos \mathcal{G} de **grafo causal** do CM (\mathcal{G}, f) .

Note pela Definição 2.24 que um CM é formalmente um modelo probabilístico em um DAG. O principal atributo de um CM que o diferencia de um modelo probabilístico genérico em um DAG é como o interpretamos. Existe uma aresta de V_1 em V_2 em um CM se e somente se V_1 é uma causa direta de V_2 .

Dentre os modelos causais, é de particular interesse o modelo linear Gaussiano.

Definição 2.25. Dizemos que (\mathcal{G}, f) é um CM linear Gaussiano de parâmetros μ e β se, existe matriz diagonal positiva, Σ , $\mu \in \mathbb{R}^{|\mathcal{V}|}$, e $\beta \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ tal que, para todo vértice V , $\beta_{V,W} = 0$ quando $W \notin Pa(V)$ e

$$V|Pa(V) \sim N \left(\mu_V + \sum_{W \in Pa(V)} \beta_{V,W} \cdot W, \Sigma_{i,i} \right)$$

O modelo causal linear Gaussiano tem algumas propriedades especiais, que tornam mais simples suas compreensão. Algumas destas são apresentadas abaixo:

Lema 2.26. Se (\mathcal{G}, f) é um CM linear Gaussiano, então \mathcal{V} segue distribuição normal multivariada.

Lema 2.27. Seja (\mathcal{G}, f) um CM linear Gaussiano com coeficientes β . Para cada $V, Y \in \mathcal{V}$, defina $\mathbb{C}_{V,Y}$ como o

conjunto de todos os caminhos direcionados de V a Y .

$$\mathbb{E}[Y] = \sum_{V \in \mathcal{V}} \sum_{C \in \mathcal{C}_{V,Y}} \mu_V \cdot \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i}$$

No próximo capítulo estudaremos consequências desta interpretação causal. Contudo, antes disso, a próxima seção desenvolverá um resultado fundamental de modelos probabilísticos em DAGs que será fundamental nos capítulos posteriores.

2.1.6. Exercícios

Exercício 2.28. Em um DAG, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, Considere que $Anc^*(\mathbb{V}) \subseteq \mathcal{V}$ é definido como o menor conjunto tal que $\mathbb{V} \subseteq Anc^*(\mathbb{V})$ e, se $V \in Anc^*(\mathbb{V})$, então $Pa(V) \subseteq Anc^*(\mathbb{V})$. Prove que $Anc(\mathbb{V}) \equiv Anc^*(\mathbb{V})$.

Exercício 2.29. Prove o [Lema 2.10](#).

Exercício 2.30. Prove que se \mathbf{Z} é ancestral, então $f(\mathbf{Z}) = \prod_{Z \in \mathbf{Z}} f(Z|Pa(Z))$.

Exercício 2.31. Sejam $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ e $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_2)$ grafos tais que $\mathcal{E}_1 \subseteq \mathcal{E}_2$. Prove que se f é compatível com \mathcal{G}_1 , então f é compatível com \mathcal{G}_2 .

Exercício 2.32. Prove o [Lema 2.14](#).

Exercício 2.33. Prove o [Lema 2.17](#).

Exercício 2.34. Prove que, para qualquer $\mathbb{V} \subseteq \mathcal{V}$, $Anc(\mathbb{V}) = Anc(Anc(\mathbb{V}))$.

Exercício 2.35. Prove que \mathbb{V} é ancestral se e somente se $Anc(\mathbb{V}) = \mathbb{V}$.

Exercício 2.36. Considere que (X_1, X_2) são independentes e tais que $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 0.5$. Além disso, $Y \equiv X_1 \cdot X_2$.

(a) Desenhe um DAG compatível com as relações de independência dadas pelo enunciado.

(b) Prove que Y e X_1 são independentes. Isso contradiz sua resposta para o item anterior?

Exercício 2.37. Para cada um dos modelos de confundidor, cadeia e colisor, dê exemplos de situações práticas em que este modelo é razoável.

Exercício 2.38. Considere que, dado T , X_1, \dots, X_n são i.i.d. e $X_i|T \sim \text{Bernoulli}(T)$. Além disso, $T \sim \text{Beta}(a, b)$.

(a) Seja $f(t, x_1, \dots, x_n)$ dada pelo enunciado. Exiba um DAG, \mathcal{G} , tal que f é compatível com \mathcal{G} .

(b) (X_1, \dots, X_n) são independentes?

(c) Determine $f(x_1, \dots, x_n)$.

Exercício 2.39. Exiba um exemplo em que V_1, V_2, V_3 sejam binárias, que V_2 seja um colisor e que, além disso, $\text{Corr}[V_1, V_3|V_2 = 1] > 0$.

Exercício 2.40. Seja $\mathcal{V} = (V_1, V_2, V_3)$ Exiba um exemplo de f sobre \mathcal{V} e grafos \mathcal{G}_1 e \mathcal{G}_2 sobre \mathcal{V} tais que $\mathcal{G}_1 \neq \mathcal{G}_2$ e f é compatível tanto com \mathcal{G}_1 quanto com \mathcal{G}_2 .

Exercício 2.41. Seja f uma densidade arbitrária sobre $\mathcal{V} = (V_1, \dots, V_n)$. Exiba um DAG sobre \mathcal{V} , \mathcal{G} , tal que f é compatível com \mathcal{G} .

Exercício 2.42. Exiba um exemplo em que V_2 é um colisor entre V_1 e V_3 , V_4 tem como único pai V_2 e V_1 e V_3 são dependentes dado V_4 .

Exercício 2.43. Prove o [Lema 2.26](#).

2.2. Independência Condicional e D-separação

Independência condicional é uma forma fundamental de indicar relações entre variáveis aleatórias. Se $\mathbf{X}_1, \dots, \mathbf{X}_d$ e \mathbf{Y} são vetores de variáveis aleatórias, definimos que $(\mathbf{X}_1, \dots, \mathbf{X}_d) | \mathbf{Y}$, isto é, $\mathbf{X}_1, \dots, \mathbf{X}_d$ são independentes dado \mathbf{Y} , se conhecido o valor de \mathbf{Y} , observar quaisquer valores de \mathbf{X} não traz informação sobre os demais valores. Nesta seção veremos que as relações de independência condicional em um CM estão diretamente ligadas ao seu grafo.

2.2.1. Independência Condicional

Definição 2.44. Dizemos que $(\mathbf{X}_1, \dots, \mathbf{X}_d)$ são independentes dado \mathbf{Y} se, para qualquer $\mathbf{x}_1, \dots, \mathbf{x}_n$ e \mathbf{y} ,

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{y}) = \prod_{i=1}^d f(\mathbf{x}_i | \mathbf{y})$$

Em particular, $(\mathbf{X}_1, \dots, \mathbf{X}_d)$ são independentes se, para quaisquer $(\mathbf{x}_1, \dots, \mathbf{x}_d)$,

$$f(\mathbf{x}_1, \dots, \mathbf{x}_d) = \prod_{i=1}^d f(\mathbf{x}_i)$$

Verificar se a [Definição 2.44](#) está satisfeita nem sempre é fácil. A princípio, ela exige obter tanto a distribuição condicional conjunta, $f(\mathbf{x}_1, \dots, \mathbf{x}_d | \mathbf{y})$, quanto cada uma das marginais, $f(\mathbf{x}_i | \mathbf{y})$. O [Lema 2.45](#) a seguir apresenta outras condições que são equivalentes a independência condicional:

Lema 2.45. *As seguintes afirmações são equivalentes:*

1. $(\mathbf{X}_1, \dots, \mathbf{X}_d)$ são independentes dado \mathbf{Y} ,
2. Existem funções, h_1, \dots, h_d tais que $f(\mathbf{x}_1, \dots, \mathbf{x}_d | \mathbf{y}) = \prod_{j=1}^d h_j(\mathbf{x}_j, \mathbf{y})$.
3. Para todo i , $f(\mathbf{x}_i | \mathbf{x}_{-i}, \mathbf{y}) = f(\mathbf{x}_i | \mathbf{y})$.
4. Para todo i , $f(\mathbf{x}_i | \mathbf{x}_1^{i-1}, \mathbf{y}) = f(\mathbf{x}_i | \mathbf{y})$.

As condições no [Lema 2.45](#) são, em geral, mais fáceis de verificar do que a definição direta de independência condicional. A seguir veremos que, em um SMC, pode ser mais fácil ainda verificar muitas das relações de independência condicional.

2.2.2. D-separação

Em um CM, é possível indicar as relações de independência incondicional em \mathcal{V} por meio do grafo associado. Intuitivamente, haverá uma dependência entre V_1 e V_2 se for possível transmitir a informação de V_1 para V_2 por um caminho que ligue ambos os vértices. Para entender se a informação pode ser transmitida por um caminho, classificaremos a seguir os vértices que o constituem.

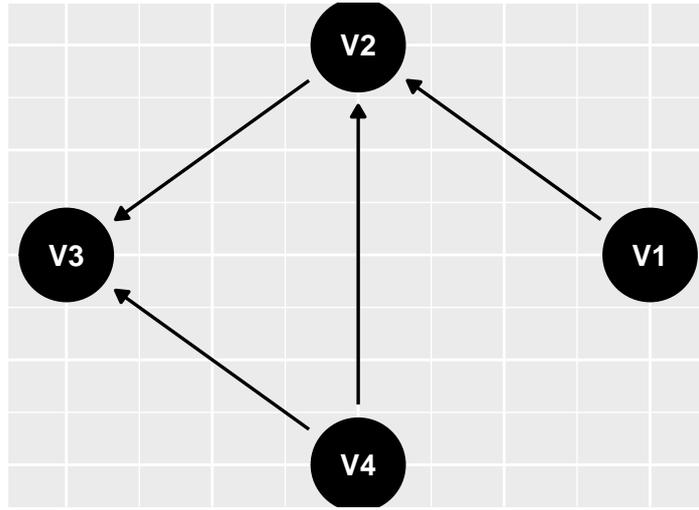


Figura 2.5.: Ilustração do conceito de bloqueio de um caminho. No caminho $(V1, V2, V4)$, $V2$ é um colisor. Isto ocorre pois, para chegar de $V1$ a $V4$ passando apenas por $V2$, as duas arestas apontam para $V2$. Já no caminho $(V1, V2, V3, V4)$ temos que $V2$ é uma cadeia. Para chegar de $V1$ a $V3$ passando por $V2$, passa-se por duas arestas, uma entrando e outra saindo de $V2$. Como $V2$ é um colisor em $(V1, V2, V4)$, este caminho está bloqueado se e somente se o valor de $V2$ é desconhecido. Como $V2$ é uma cadeia em $(V1, V2, V3, V4)$, esse caminho está bloqueado quando o valor de $V2$ é conhecido.

Definição 2.46. Seja $C = (C_1, \dots, C_n)$ um caminho em um DAG, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Para cada $2 \leq i \leq n - 1$:

- C_i é um **colisor** em C se $(C_{i-1}, C_i) \in \mathcal{E}$ e $(C_{i+1}, C_i) \in \mathcal{E}$, isto é, existem arestas apontando de C_{i-1} e de C_{i+1} para C_i . Neste caso, desenhamos $C_{i-1} \rightarrow C_i \leftarrow C_{i+1}$.

Note que a classificação na [Definição 2.46](#) generaliza os exemplos de DAG's com 3 vértices na [Seção 2.1.4](#).

Essa classificação é ilustrada com o DAG na [figur 2.5](#). Existem dois caminhos que vão de V_1 a V_4 : $V_1 \rightarrow V_2 \leftarrow V_4$ e $V_1 \rightarrow V_2 \rightarrow V_3 \leftarrow V_4$. No primeiro caminho V_2 é um colisor, pois o caminho passa por duas arestas que apontam para V_2 . Já no segundo caminho V_2 é uma cadeia e V_3 é um colisor. Note que a classificação do vértice depende do caminho analisado. Enquanto que no primeiro caminho V_2 é um colisor, no segundo V_2 é uma cadeia

Com base nas conclusões da [Seção 2.1.4](#), é possível compreender a racionalidade da [Definição 2.46](#). Na [Seção 2.1.4](#) vimos que, se Z não é um colisor entre X e Y , então X e Y são independentes dado Z . Por analogia, podemos intuir que um vértice que não é um colisor num caminho não permite a passagem de informação quando seu valor é conhecido. Similarmente, na [Seção 2.1.4](#), se Z é um colisor entre X e Y , então X e Y são independentes. Assim, também podemos intuir que um vértice que é um colisor em um caminho não permite a passagem de informação quando seu valor e o de seus descendentes é desconhecido. Finalmente, a informação não passa pelo caminho quando ela não passa por pelo menos um de seus vértices. Neste caso, dizemos que o caminho está *bloqueado*:

Definição 2.47. Seja $C = (C_1, \dots, C_n)$ um caminho em um DAG, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Dizemos que C está bloqueado dado $\mathbf{Z} \subset \mathcal{V}$, se

1. Existe algum $2 \leq i \leq n - 1$ tal que C_i não é um colisor em C e $C_i \in \mathbf{Z}$, ou
2. Existe algum $2 \leq i \leq n - 1$ tal que C_i é um colisor em C e $C_i \notin \text{Anc}(\mathbf{Z})$.

Finalmente, dizemos que \mathbb{V}_1 está d-separado de \mathbb{V}_2 dado \mathbb{V}_3 se todos os caminhos de \mathbb{V}_1 a \mathbb{V}_2 estão bloqueados dado \mathbb{V}_3 :

Definição 2.48. Seja $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ um DAG. Para $V_1, V_2, V_3 \subseteq \mathcal{V}$, dizemos que V_1 está d-separado de V_2 dado V_3 se, para todo caminho $C = (C_1, \dots, C_n)$ tal que $C_1 \in V_1$ e $C_n \in V_2$, C está bloqueado dado V_3 . Neste caso, escrevemos $V_1 \perp V_2 | V_3$.

Intuitivamente, se $V_1 \perp V_2 | V_3$, então não é possível passar informação de V_1 a V_2 quando V_3 é conhecido. Assim, temos razão para acreditar que V_1 é condicionalmente independente de V_2 dado V_3 , isto é $V_1 \perp V_2 | V_3$. Esta conclusão é apresentada no [Teorema 2.49](#) a seguir:

Teorema 2.49. *Seja $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ um DAG e \mathcal{V} um conjunto de variáveis aleatórias. V_1 está d-separado de V_2 dado V_3 se e somente se, para todo f compatível com \mathcal{G} , $V_1 \perp^f V_2 | V_3$.*

Exemplo 2.50. Considere o DAG na [figur 2.5](#). Para avaliar se V_1 e V_3 são d-separados, precisamos analisar todos os caminhos de um para o outro. Estes caminhos são: $V_1 \rightarrow V_2 \rightarrow V_3$, e $V_1 \rightarrow V_2 \leftarrow V_4 \rightarrow V_3$. No primeiro caminho V_2 não é um colisor e, assim, o caminho não está bloqueado marginalmente. Portanto, V_1 e V_3 não são d-separados marginalmente. Por outro lado, no segundo caminho V_2 é um colisor e V_4 não o é. Assim, condicionando em V_2 , este caminho não está bloqueado. Portanto, V_1 e V_3 não são d-separados dado V_2 . Finalmente, dado V_2 e V_4 , ambos os caminhos estão bloqueados, pois V_2 não é um colisor no primeiro e V_4 não é um colisor no segundo. Assim, V_1 e V_3 são d-separados dado (V_2, V_4) . Para treinar este raciocínio, continue analisando a d-separação entre V_1 e V_4 .

O algoritmo para testar d-separação está implementado em diversos pacotes. Além disso, é possível utilizar o [Teorema 2.49](#) para enunciar todas as relações de independência condicional que são necessárias em um grafo. Estas implementações estão ilustradas abaixo:

```
# Especificar o grafo
grafo <- "dag{
  V1 -> V2 <- V4;
  V2 -> V3 <- V4
}"

dseparated(grafo, "V1", "V3", c("V2"))

## [1] FALSE

dseparated(grafo, "V1", "V3", c("V4"))

## [1] FALSE

dseparated(grafo, "V1", "V3", c("V2", "V4"))

## [1] TRUE

impliedConditionalIndependencies(grafo)

## V1 _||_ V3 | V2, V4
## V1 _||_ V4
```

Exemplo 2.51. Considere que \mathbb{V}_1 e \mathbb{V}_2 não são d-separados dado \mathbb{V}_3 . O Teorema 2.49 garante apenas que existe algum f compatível com o DAG tal que \mathbb{V}_1 e \mathbb{V}_2 são condicionalmente dependentes dado \mathbb{V}_3 segundo f . É possível mostrar que o conjunto de f 's compatíveis com o grafo em que \mathbb{V}_1 e \mathbb{V}_2 são condicionalmente independentes dado \mathbb{V}_3 é relativamente pequeno àquele em que \mathbb{V}_1 e \mathbb{V}_2 são condicionalmente dependentes. Estudaremos um caso em que é possível observar esta relação em mais detalhe.

Considere que V_1, V_2 , e Z são binárias e formam o grafo $V_1 \leftarrow Z \rightarrow V_2$, isto é, Z é um confundidor. Além disso, $\mathbb{P}(Z = 1) = 0.5$, $\mathbb{P}(V_i = 1|Z = j) =: p_j$. Como V_3 é um confundidor, V_1 e V_2 não são d-separados marginalmente. Para quais valores de p temos que V_1 e V_2 são marginalmente independentes? Para que V_1 e V_2 sejam independentes, é necessário que $Cov[V_1, V_2] = 0$. Note que

$$\begin{aligned}\mathbb{E}[V_i] &= \mathbb{E}[\mathbb{E}[V_i|Z]] = 0.5p_1 + 0.5p_0 \\ \mathbb{E}[V_1 V_2] &= \mathbb{E}[\mathbb{E}[V_1 V_2|Z]] = 0.5p_1^2 + 0.5p_0^2\end{aligned}$$

Assim, para que $Cov[V_1, V_2] = 0$, temos:

$$\begin{aligned}0.5p_1^2 + 0.5p_0^2 &= (0.5p_1 + 0.5p_0)(0.5p_1 + 0.5p_0) \\ 0.5p_1^2 + 0.5p_0^2 &= 0.25p_1^2 + 0.5p_1p_0 + 0.25p_0^2 \\ 0.25p_1^2 - 0.5p_1p_0 + 0.25p_0^2 &= 0 \\ 0.25(p_1 - p_0)^2 &= 0 \\ p_1 &= p_0\end{aligned}$$

Em outras palavras, dentre todos (p_0, p_1) no quadrado $[0, 1]^2$, somente os valores no segmento $p_1 = p_0$ tem alguma chance de levarem à independência entre V_1 e V_2 . Se imaginarmos que (p_0, p_1) são equidistribuídos em $[0, 1]^2$, então a probabilidade de sortearmos valores em que V_1 e V_2 são independentes é 0.

Em conclusão, como V_1 e V_2 não são d-separados, somente para um conjunto pequeno de possíveis f 's temos que V_1 e V_2 são independentes.

2.2.3. Exercícios

Exercício 2.52. Considere que f é uma densidade sobre $\mathcal{V} = (V_1, V_2, V_3, V_4)$ que é compatível com o grafo em [figura 2.6](#). Além disso, cada $V_i \in \{0, 1\}$, $V_1, V_2 \sim \text{Bernoulli}(0.5)$, $V_3 \equiv V_1 \cdot V_2$ e $\mathbb{P}(V_4 = i|V_3 = i) = 0.9$, para todo i .

- V_1 e V_2 são d-separados dado V_3 ?
- V_1 e V_2 são condicionalmente independentes dado V_3 ?
- V_1 e V_2 são d-separados dado V_4 ?
- V_1 e V_2 são condicionalmente independentes dado V_4 ?

Exercício 2.53. Prove que se um caminho, $C = (C_1, \dots, C_n)$, está bloqueado dado \mathbb{V} , então sempre que C é um sub-caminho de C^* , isto é, $C^* = (A_1, \dots, A_m, C_1, \dots, C_n, B_1, \dots, B_l)$, temos que C^* está bloqueado dado \mathbb{V} .

Exercício 2.54. Prove que se $\mathbb{V}_1 \perp \mathbb{V}_3|\mathbb{V}_4$ e $\mathbb{V}_2 \perp \mathbb{V}_3|\mathbb{V}_4$, então $\mathbb{V}_1 \cup \mathbb{V}_2 \perp \mathbb{V}_3|\mathbb{V}_4$.

Exercício 2.55. Prove que se $\mathbb{V}_1 \perp \mathbb{V}_2|\mathbb{V}_3$, então para todo $V \in (\text{Anc}(\mathbb{V}_3) - \mathbb{V}_3)$, $V \perp \mathbb{V}_1|\mathbb{V}_3$ ou $V \perp \mathbb{V}_2|\mathbb{V}_3$.

Exercício 2.56. Sejam $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ e $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_2)$ grafos tais que $\mathcal{E}_1 \subseteq \mathcal{E}_2$. Prove que se $\mathbb{V}_1 \perp^d \mathbb{V}_2|\mathbb{V}_3$ em \mathcal{G}_2 , então $\mathbb{V}_1 \perp^d \mathbb{V}_2|\mathbb{V}_3$ em \mathcal{G}_1 .

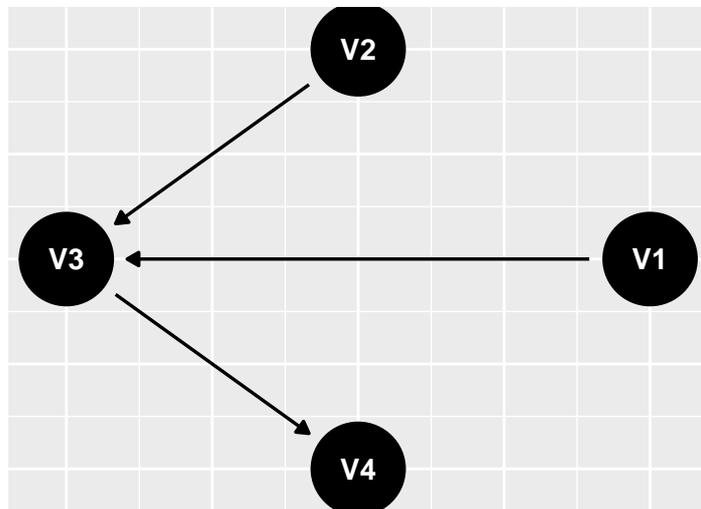


Figura 2.6.: Exemplo em que V4 é um descendente de um colisor, V3.

3. Intervenções

3.1. O modelo de probabilidade para intervenções

Com base no modelo estrutural causal discutido no [capítulo 2](#), agora estabeleceremos um significado para o efeito causal de uma variável em outra.

Para iniciar esta discussão, considere as variáveis Z (Sexo), X (Tratamento), e Y (Cura), discutidas no [capítulo 1](#). Podemos considerar que Z é uma causa tanto de X quanto de Y e que X é uma causa de Y . Assim, podemos representar as relações causais entre estas variáveis por meio do grafo na [figur 3.1](#). Usando este grafo, podemos discutir mais a fundo porque a probabilidade condicional de cura dado tratamento é distinta do efeito causal do tratamento na cura.

Quando calculamos a probabilidade condicional de cura dado o tratamento, estamos perguntando: “Qual é a probabilidade de que um indivíduo selecionado aleatoriamente da população se cure dado que **aprendemos** que recebeu o tratamento?” Para responder a esta pergunta, propagamos a informação do tratamento usado em todos os caminhos do tratamento para a cura. Assim, além do efeito direto que o tratamento tem na cura, o tratamento também está associado ao sexo do paciente, o que indiretamente traz mais informação sobre a cura deste. Isto é, neste caso o tratamento traz informação tanto sobre seus efeitos (cura), quanto sobre suas causas (sexo). Uma outra maneira de verificar estas afirmações é calculando diretamente $f(y|x)$:

$$\begin{aligned} f(y|x) &= \sum_s f(z, y|x) \\ &= \sum_s \frac{f(z, y, x)}{f(x)} \\ &= \sum_s \frac{f(z, x)f(y|z, x)}{f(x)} \\ &= \sum_s f(z|x)f(y|z, x) \end{aligned} \tag{3.1}$$

Notamos na [likning \(3.1\)](#) que $f(y|x)$ é a média das probabilidades de cura em cada sexo, $f(y|z, x)$, ponderadas pela distribuição do sexo após aprender o tratamento do indivíduo, $f(z|x)$.

A probabilidade condicional de cura dado tratamento não corresponde àquilo que entendemos por efeito causal de tratamento em cura. Este efeito é a resposta para a pergunta: “Qual a probabilidade de que um indivíduo selecionado aleatoriamente da população se cure dado que **prescrevemos** a ele o tratamento?”. Ao contrário da primeira pergunta, em que apenas **observamos** a população, nesta segunda fazemos uma **intervenção** sobre o comportamento do indivíduo. Assim, estamos fazendo uma pergunta sobre uma distribuição de probabilidade diferente, em que estamos agindo sobre a unidade amostral. Por exemplo, suponha que prescreveríamos o tratamento a qualquer indivíduo que fosse amostrado. Neste caso, saber qual tratamento foi aplicado não traria qualquer informação sobre o sexo do indivíduo. Em outras palavras, se chamarmos $f(y|do(x))$ como a probabilidade de

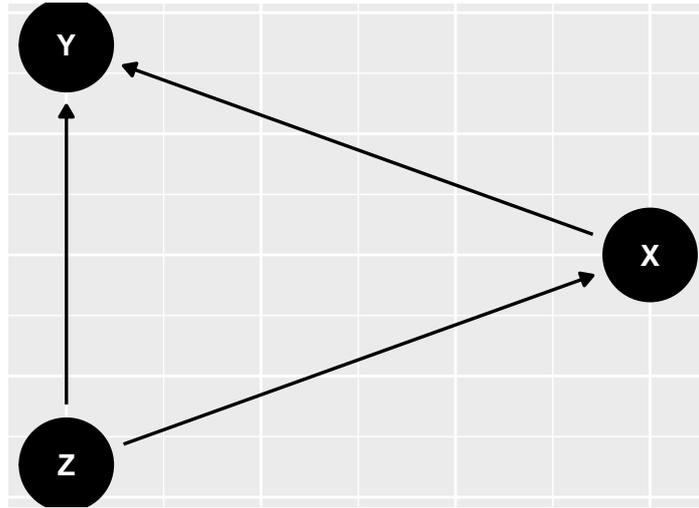


Figura 3.1.: Grafo que representa as relações causais entre Z (Sexo), X (Tratamento), e Y (Cura).

cura dado que fazemos uma intervenção no tratamento, faria sentido obtermos:

$$f(y|do(x)) = \sum_s f(z) f(y|z, x) \quad (3.2)$$

Na [likning \(3.2\)](#) temos que o efeito causal do tratamento na cura é a média ponderada das probabilidades de cura em cada sexo ponderada pelas probabilidades de sexo de um indivíduo retirado aleatoriamente da população. Isto é, ao contrário da [likning \(3.1\)](#), a distribuição do sexo do indivíduo não é alterada quando fazemos uma intervenção sobre o tratamento.

Com base neste exemplo, podemos generalizar o que entendemos por intervenção. Quando fazemos uma intervenção em uma variável, V_1 , tomamos uma ação para que V_1 assuma um determinado valor. Assim, as demais variáveis que comumente seriam causas de V_1 deixam de sê-lo. Por exemplo, para o caso na [figur 3.1](#), o modelo de intervenção removeria a aresta de Sexo para Tratamento, resultado na [figur 3.2](#).

Com base nas observações acima, finalmente podemos definir o modelo de probabilidade sob intervenção:

Definição 3.1. Seja $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ um DAG, (\mathcal{G}, f) um CM ([Definição 2.24](#)), e $\mathbb{V}_1 \subseteq \mathcal{V}$. O modelo de probabilidade obtido após uma intervenção em \mathbb{V}_1 é dado por:

$$f(\mathcal{V}|do(\mathbb{V}_1)) := \prod_{V_2 \in \mathbb{V}_2} f(V_2|Pa(V_2)) \quad , \text{ ou equivalentemente}$$

$$f(\mathcal{V}|do(\mathbb{V}_1 = \mathbf{v}_1)) := \left(\prod_{(v_1, V_1) \in (\mathbf{v}_1, \mathbb{V}_1)} \mathbb{I}(V_1 = v_1) \right) \cdot \left(\prod_{V_2 \notin \mathbb{V}_1} f(V_2|Pa(V_2)) \right)$$

Para compreender a [Definição 3.1](#), podemos comparar o modelo de intervenção com o modelo observacional:

$$f(\mathbb{V}_2|\mathbb{V}_1) \propto f(\mathbb{V}_1, \mathbb{V}_2) = \left(\prod_{V_1 \in \mathbb{V}_1} f(V_1|Pa(V_1)) \right) \cdot \left(\prod_{V_2 \in \mathbb{V}_2} f(V_2|Pa(V_2)) \right)$$

No modelo observacional, a densidade de \mathbb{V}_2 dado \mathbb{V}_1 é proporcional ao produto, para todos os vértices, da

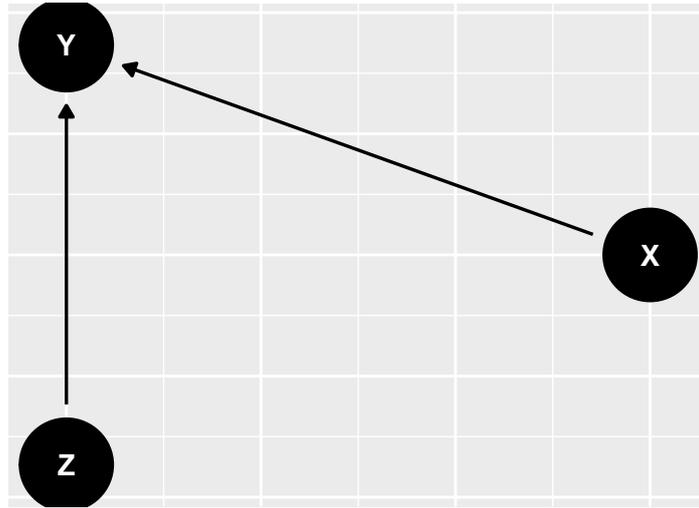


Figura 3.2.: Grafo que representa as relações causais entre S (Sexo), T (Tratamento), e C (Cura) quando há uma intervenção sobre T.

densidade do vértice dadas suas causas. Ao contrário, no modelo de intervenção supomos que os vértices em \mathbb{V}_1 são pré-fixados e, assim, não são gerados por suas causas usuais. Assim, na [Definição 3.1](#), a densidade de \mathbb{V}_2 dada uma intervenção em \mathbb{V}_1 é dada o produto somente nos vértices de \mathbb{V}_2 das densidades do vértice dadas suas causas.

Esta análise é formalizada no [Lema 3.2](#):

Lema 3.2. *Seja $\mathcal{G}(\bar{\mathbf{X}})$ o grafo obtido retirando-se de \mathcal{G} todas as arestas que apontam para algum vértice em \mathbf{X} . A densidade $f^* \equiv f(\mathcal{V}|\text{do}(\mathbf{X} = \mathbf{x}))$ é compatível com $\mathcal{G}(\bar{\mathbf{X}})$. Além disso, \mathbf{X} é degenerada em \mathbf{x} segundo f^* .*

Com base na discussão acima, podemos definir o **efeito causal** que um conjunto de variáveis, \mathbf{X} , tem em outro conjunto, \mathbf{Y} :

Definição 3.3. $\mathbb{E}[\mathbf{Y}|\text{do}(\mathbf{X})] := \int \mathbf{y} \cdot f(\mathbf{y}|\text{do}(\mathbf{X}))d\mathbf{y}$.

Definição 3.4. O efeito causal médio, $ACE_{X,Y}$,¹ de $X \in \mathfrak{R}$ em $Y \in \mathfrak{R}$ é dado por:

$$ACE_{X,Y} = \begin{cases} \mathbb{E}[Y|\text{do}(X = 1)] - \mathbb{E}[Y|\text{do}(X = 0)] & , \text{ se } X \text{ é binário,} \\ \frac{d\mathbb{E}[Y|\text{do}(X=x)]}{dx} & , \text{ se } X \text{ é contínuo.} \end{cases}$$

Quando não há ambiguidade, escrevemos simplesmente ACE ao invés de $ACE_{X,Y}$.

Com a [Definição 3.4](#) podemos finalmente desvendar o Paradoxo de Simpson discutido no [capítulo 1](#). Veremos que o método que desenvolvemos resolve a questão com simplicidade, assim trazendo clareza ao Paradoxo.

Exemplo 3.5. Considere que $(X, Y, Z) \in \mathfrak{R}^3$ são tais que X e Y são as indicadores de que, respectivamente, o paciente recebeu o tratamento e se curou. Além disso, suponha que a distribuição conjunta de (X, Y, Z) é dada

¹A sigla ACE tem como origem a expressão em inglês, *Average Causal Effect*. Optamos por manter a sigla sem tradução para facilitar a comparação com artigos da área. Em outros contextos, este termo também é chamado de *Average Treatment Effect* e recebe o acrônimo ATE.

pelas frequências na [tabela 1.1](#). Isto é:

$$\begin{aligned}\mathbb{P}(Z = 1) &= \frac{25 + 55 + 71 + 192}{700} \approx 0.49 \\ \mathbb{P}(Z = 0) &= 1 - \mathbb{P}(Z = 1) \approx 0.51 \\ \mathbb{P}(Z = 1|X = 0) &= \frac{25 + 55}{25 + 55 + 36 + 234} \approx 0.23 \\ \mathbb{P}(Z = 1|X = 1) &= \frac{71 + 192}{71 + 192 + 6 + 81} \approx 0.75 \\ \mathbb{P}(Y = 1|X = 0, Z = 0) &= \frac{234}{234 + 36} \approx 0.87 \\ \mathbb{P}(Y = 1|X = 1, Z = 0) &= \frac{81}{81 + 6} \approx 0.93 \\ \mathbb{P}(Y = 1|X = 0, Z = 1) &= \frac{55}{25 + 55} \approx 0.69 \\ \mathbb{P}(Y = 1|X = 1, Z = 1) &= \frac{192}{71 + 192} \approx 0.73\end{aligned}$$

Agora, veremos que a probabilidade de Y dada uma intervenção em X depende do DAG usado no modelo causal estrutural.

Suponha que Z é a indicadora de que o sexo do paciente é masculino. Neste caso, utilizaremos como grafo causal aquele em [figur 3.1](#). este grafo, obtemos:

$$\mathbb{P}_1(Y = i, Z = j|do(X = k)) = \mathbb{P}(Z = j)\mathbb{P}(Y = i|X = k, Z = j) \quad \text{Definição 3.1} \quad (3.3)$$

Assim,

$$\begin{aligned}\mathbb{P}_1(Y = 1|do(X = 1)) &= \mathbb{P}_1(Y = 1, Z = 0|do(X = 1)) + \mathbb{P}_1(Y = 1, Z = 1|do(X = 1)) \\ &= \mathbb{P}(Z = 0)\mathbb{P}(Y = 1|X = 1, Z = 0) + \mathbb{P}(Z = 1)\mathbb{P}(Y = 1|X = 1, Z = 1) \quad \text{likning (3.3)} \\ &\approx 0.51 \cdot 0.93 + 0.49 \cdot 0.73 \approx 0.83\end{aligned}$$

$$\begin{aligned}\mathbb{P}_1(Y = 1|do(X = 0)) &= \mathbb{P}_1(Y = 1, Z = 0|do(X = 0)) + \mathbb{P}_1(Y = 1, Z = 1|do(X = 0)) \\ &= \mathbb{P}(Z = 0)\mathbb{P}(Y = 1|X = 0, Z = 0) + \mathbb{P}(Z = 1)\mathbb{P}(Y = 1|X = 0, Z = 1) \quad \text{likning (3.3)} \\ &\approx 0.51 \cdot 0.87 + 0.49 \cdot 0.69 \approx 0.78\end{aligned}$$

Portanto, o efeito causal do tratamento na cura quando Z é o sexo do paciente é obtido abaixo:

$$\begin{aligned}ACE_1 &= \mathbb{E}_1[Y|do(X = 1)] - \mathbb{E}_1[Y|do(X = 0)] \quad \text{Definição 3.4} \\ &= \mathbb{P}_1(Y = 1|do(X = 1)) - \mathbb{P}_1(Y = 1|do(X = 0)) \approx 0.05 \quad \text{Definição 3.3}\end{aligned}$$

Como esperado da discussão na [Seção 1.1](#), o tratamento tem efeito causal médio positivo, isto é, ele aumenta a probabilidade de cura do paciente.

A seguir, consideramos que Z é a indicadora de pressão sanguínea elevada do paciente. Assim, tomamos o grafo causal como aquele na [figur 3.3](#). Utilizando este grafo, obtemos:

$$\mathbb{P}_2(Y = i, Z = j|do(X = k)) = \mathbb{P}(Z = j|X = k)\mathbb{P}_1(Y = i|X = k, Z = j) \quad \text{Definição 3.1} \quad (3.4)$$

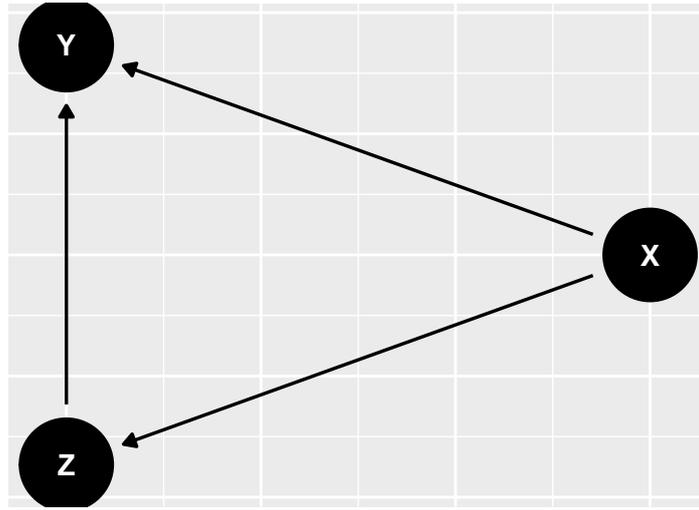


Figura 3.3.: Grafo que representa as relações causais entre Z (Pressão sanguínea elevada), X (Tratamento), e Y (Cura).

Assim,

$$\begin{aligned} \mathbb{P}_2(Y = 1|do(X = 1)) &= \mathbb{P}_2(Y = 1, Z = 0|do(X = 1)) + \mathbb{P}_2(Y = 1, Z = 1|do(X = 1)) \\ &= \mathbb{P}(Z = 0|X = 1)\mathbb{P}(Y = 1|X = 1, Z = 0) + \mathbb{P}(Z = 1|X = 1)\mathbb{P}(Y = 1|X = 1, Z = 1) \quad \text{likning (3.4)} \\ &\approx 0.25 \cdot 0.93 + 0.75 \cdot 0.73 \approx 0.78 \end{aligned}$$

$$\begin{aligned} \mathbb{P}_2(Y = 1|do(X = 0)) &= \mathbb{P}_2(Y = 1, Z = 0|do(X = 0)) + \mathbb{P}_2(Y = 1, Z = 1|do(X = 0)) \\ &= \mathbb{P}(Z = 0|X = 0)\mathbb{P}(Y = 1|X = 0, Z = 0) + \mathbb{P}(Z = 1|X = 0)\mathbb{P}(Y = 1|X = 0, Z = 1) \quad \text{likning (3.4)} \\ &\approx 0.77 \cdot 0.87 + 0.23 \cdot 0.69 \approx 0.83 \end{aligned}$$

Portanto, o efeito causal do tratamento na cura quando Z é a pressão sanguínea do paciente é obtido abaixo:

$$\begin{aligned} ACE_1 &= \mathbb{E}_2[Y|do(X = 1)] - \mathbb{E}_2[Y|do(X = 0)] && \text{Definição 3.4} \\ &= \mathbb{P}_2(Y = 1|do(X = 1)) - \mathbb{P}_2(Y = 1|do(X = 0)) \approx -0.05 && \text{Definição 3.3} \end{aligned}$$

Como esperado da discussão na [Seção 1.1](#), o tratamento tem efeito causal médio negativo, isto é, ele tem como efeito colateral grave a elevação da pressão sanguínea do paciente, reduzindo a probabilidade de cura deste.

Comparando as expressões obtidas em ACE_1 e ACE_2 , verificamos que o grafo causal desempenha papel fundamental na determinação do modelo de probabilidade sob intervenção. Ademais, o uso do grafo causal adequado em cada situação formaliza a discussão qualitativa desenvolvida na [Seção 1.1](#). Não há paradoxo!

Se (\mathcal{G}, f) é um CM linear Gaussiano, então é possível obter uma equação direta para o ACE . Este resultado é apresentado no [Teorema 3.6](#) abaixo.

Teorema 3.6. *Se (\mathcal{G}, f) é um CM linear Gaussiano de parâmetros μ e β e $\mathbb{C}_{X,Y}$ é o conjunto de todos os caminhos*

direcionados de X a Y , então

$$ACE_{X,Y} = \sum_{C \in \mathcal{C}_{X,Y}} \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i}.$$

O Teorema 3.6 indica um algoritmo para calcular o $ACE_{X,Y}$ em um CM linear Gaussiano. Primeiramente, para cada caminho direcionado de X em Y calcula-se o produto dos coeficientes de regressão ligados a este caminho. Se imaginarmos os vértices no meio do caminho como mediadores, então estamos combinando o efeito de X em C_2 , de C_2 em $C_3 \dots$ e de C_{m-1} em Y para obter o efeito total de X em Y por este caminho. Ao final, somamos os efeitos totais obtidos por todos os caminhos. Cada caminho direcionado indica uma forma em que X pode ter efeito sobre Y . Ao levarmos todas as formas em consideração, obtemos o efeito causal médio.

Além do efeito causal médio, às vezes desejamos determinar o efeito causal de X em Y quando observamos que a unidade amostral faz parte de determinado estrato da população. Em outras palavras, desejamos saber o efeito causal de X em Y quando observamos que outras variáveis, \mathbf{Z} , assumem um determinado valor.

Definição 3.7. O efeito causal médio condicional, CACE, de $X \in \mathfrak{X}$ em $Y \in \mathfrak{Y}$ dado \mathbf{Z} é:

$$CACE(\mathbf{Z}) = \begin{cases} \mathbb{E}[Y|do(X=1), \mathbf{Z}] - \mathbb{E}[Y|do(X=0), \mathbf{Z}] & , \text{ se } X \text{ é binário,} \\ \frac{d\mathbb{E}[Y|do(X=x), \mathbf{Z}]}{dx} & , \text{ se } X \text{ é contínuo.} \end{cases}$$

Uma vez estabelecido o modelo de probabilidade utilizado quando estudamos intervenções, agora podemos fazer inferência sobre o efeito causal. Para realizar tal inferência, em geral teremos de abordar duas questões:

1. **Identificação causal:** Temos acesso a dados que são gerados segundo a distribuição observacional. Como é possível determinar o efeito causal em termos da distribuição observacional?
2. **Estimação:** Uma vez estabelecida uma ligação entre a distribuição observacional dos dados e o efeito causal, como é possível estimá-lo?

Nas próximas seções estudaremos algumas estratégias gerais para a resolução destas questões. Consideraremos que desejamos medir o efeito causal de X em Y , onde $X, Y \in \mathcal{V}$.

3.1.1. Exercícios

Exercício 3.8. Considere que X_1 e X_2 são variáveis binárias. Também considere as seguintes definições: **ACE** := $\mathbb{P}(X_2 = 1|do(X_1 = 1)) - \mathbb{P}(X_2 = 1|do(X_1 = 0))$, e **RD** := $\mathbb{P}(X_2 = 1|X_1 = 1) - \mathbb{P}(X_2 = 1|X_1 = 0)$. Explique em palavras a diferença entre ACE e RD e apresente um exemplo em que essa diferença ocorre.

Exercício 3.9 (?[p.32]). (X_1, X_2, X_3, X_4) são variáveis binárias tais que X_{i-1} é a única causa imediata de X_i . Além disso, $\mathbb{P}(X_1 = 1) = 0.5$, $\mathbb{P}(X_i = 1|X_{i-1} = 1) = p_{11}$ e $\mathbb{P}(X_i = 1|X_{i-1} = 0) = p_{01}$. Calcule:

- (a) $\mathbb{P}(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0)$,
- (b) $\mathbb{P}(X_4 = 1|X_1 = 1)$, $\mathbb{P}(X_4 = 1|do(X_1 = 1))$,
- (c) $\mathbb{P}(X_1 = 1|X_4 = 1)$, $\mathbb{P}(X_1 = 1|do(X_4 = 1))$, e
- (d) $\mathbb{P}(X_3 = 1|X_1 = 0, X_4 = 1)$

Exercício 3.10 ([p.29]). Considere que (U_1, U_2, U_3) são independentes e tais que $U_i \sim N(0, 1)$. Também, $X_1 \equiv U_1$, $X_2 \equiv 3^{-1}X_1 + U_2$, e $X_3 \equiv 2^{-4}X_2 + U_3$. Considere que X_1 é a causa imediata de X_2 , que por sua vez é a causa imediata de X_3 . Além disso, cada U_i influencia diretamente somente X_i .

- (a) Desenhe o DAG que representa a estrutura causal indicada no enunciado.
- (b) Calcule $\mathbb{E}[X_2|X_1 = 3]$ e $\mathbb{E}[X_2|do(X_1 = 3)]$.
- (c) Calcule $\mathbb{E}[X_3|X_1 = 6]$ e $\mathbb{E}[X_3|do(X_1 = 6)]$.
- (d) Calcule $\mathbb{E}[X_1|X_2 = 1]$ e $\mathbb{E}[X_1|do(X_2 = 1)]$.
- (e) Calcule $\mathbb{E}[X_2|X_1 = 1, X_3 = 3]$, $\mathbb{E}[X_2|X_1 = 1, do(X_3 = 3)]$, e $\mathbb{E}[X_2|do(X_1 = 1), X_3 = 3]$.

3.2. Controlando confundidores (critério backdoor)

Um confundidor é uma causa comum, direta ou indireta, de X em Y . Na existência de confundidores, a regressão de Y em X no modelo observacional, $\mathbb{E}[Y|X]$, é diferente desta regressão no modelo de intervenção, $\mathbb{E}[Y|do(X)]$. Isto ocorre pois, quando calculamos $\mathbb{E}[Y|X]$, utilizamos toda a informação em X para prever Y . Esta informação inclui não apenas o efeito causal de X em Y , como também a informação que X traz indiretamente sobre Y pelo fato de ambas estarem associados aos seus confundidores.

Para ilustrar este raciocínio, podemos revisitar o [Exemplo 3.5](#). uma vez que Sexo (Z) é causa comum do Tratamento (X) e da Cura (Y), Z é um confundidor. Quando calculamos $f(y|x)$ ([likning \(3.1\)](#)), utilizamos não só o efeito direto de X em Y , expresso em $f(y|x, z)$, como também a informação que indireta que X traz sobre Y por meio do confundidor Z , expressa pela combinação de $f(z|x)$ com $f(y|x, z)$.

Esta seção desenvolve uma estratégia para medir o efeito causal chamada de critério *backdoor*, que consiste em bloquear todos os caminhos de informação que passam por confundidores:

Definição 3.11. Seja $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ um grafo causal e $X, Y \in \mathcal{V}$. Dizemos que $\mathbf{Z} \subseteq \mathcal{V} - \{X, Y\}$ satisfaz o critério “backdoor” se:

1. $X \notin Anc(\mathbf{Z})$,
2. Para todo caminho de X em Y , $C = (X, C_2, \dots, C_{n-1}, Y)$ tal que $(C_2, X) \in \mathcal{E}$, C está bloqueado dado \mathbf{Z} .

Exemplo 3.12. No [Exemplo 3.5](#) o único caminho de X em Y em que o vértice ligado a X é pai de X é $X \leftarrow Z \rightarrow Y$. Como Z é um confundidor neste caminho, ele o bloqueia. Assim, Z satisfaz o critério backdoor.

Exemplo 3.13. Considere o grafo causal na [figur 3.4](#). Para aplicar o critério backdoor, devemos identificar todos os caminhos de X em Y em que o vértice ligado a X é pai de X , isto é, temos $X \leftarrow$. O único caminho deste tipo é: $X \leftarrow Z \leftarrow W \rightarrow Y$. Neste caminho, Z é uma cadeia e W é um confundidor. Assim, é possível bloquear este caminho condicionando em Z , em W , e em (Z, W) . Isto é, todas estas combinações satisfazem o critério backdoor.

Exemplo 3.14. Considere o grafo causal na [figur 3.5](#). Para aplicar o critério backdoor, encontramos todos os caminhos de X em Y em que o vértice ligado a X é pai de X . Há dois caminhos deste tipo: $X \leftarrow A \rightarrow B \rightarrow Y$ e $X \leftarrow C \rightarrow Y$. Como A e C são confundidores, respectivamente, no primeiro e segundo caminhos, (A, C) bloqueia ambos eles. Assim (A, C) satisfaz o critério backdoor. Você consegue encontrar outro conjunto de variáveis que satisfaz o critério backdoor?

Também é possível identificar os conjuntos de variáveis que satisfazem o critério backdoor por meio do pacote *dagitty*, como ilustrado a seguir:

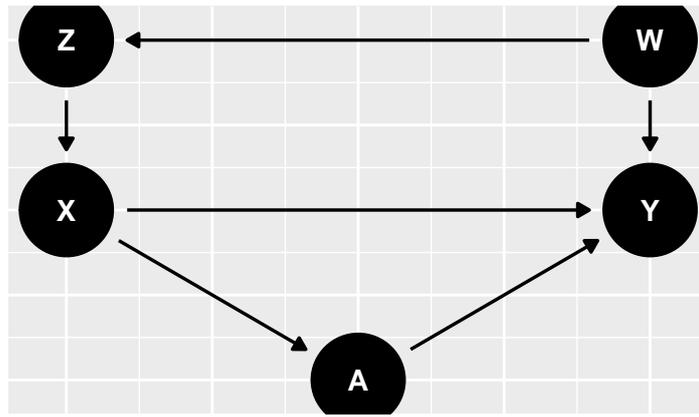


Figura 3.4.: Para medir o efeito causal de X em Y , podemos aplicar o critério backdoor. Neste grafo o único caminho aplicável ao critério backdoor é (X, Z, W, Y) . Neste caminho, Z é uma cadeia e W é um confundidor. Assim, todas as possibilidades dentre $Z, W, (Z, W)$ bloqueiam o caminho e satisfazem o critério backdoor.

```

library(dagitty)
# Especificar o grafo
grafo <- dagitty("dag{
  X[e] Y[o]
  A -> { X B }; B -> { Y }; C -> { X Y };
  X -> { D Y }; D -> Y}")

adjustmentSets(grafo, type = "all")

## { A, C }
## { B, C }
## { A, B, C }

```

O critério backdoor generaliza duas condições especiais que são muito utilizadas. Em uma primeira condição, o valor de X é gerado integralmente por um aleatorizador, independente de todas as demais variáveis. Esta ideia é captada pela [Definição 3.15](#), abaixo:

Definição 3.15. Dizemos que X é um experimento aleatorizado simples se X é ancestral.

Em um experimento aleatorizado simples não há confundidores. Assim, \emptyset satisfaz o critério backdoor:

Lema 3.16. Se X é um experimento aleatorizado simples, então \emptyset satisfaz o critério backdoor.

Veremos que em um experimento aleatorizado simples a distribuição intervencional é igual à distribuição observacional. Assim, $\mathbb{E}[Y|do(X)] = \mathbb{E}[Y|X]$ e a inferência causal é reduzida à inferência comumente usadas para a distribuição observacional.

Além disso, o conjunto de todos os pais de X também satisfaz o critério backdoor:

Lema 3.17. $\mathbf{Z} = Pa(X)$ satisfaz o critério backdoor para medir o efeito causal de X em Y .

A seguir, veremos como o critério backdoor permite a identificação causal, isto é, uma equivalência entre quantidades de interesse obtidas pelo modelo de intervenção e quantidades obtidas pelo modelo observacional.

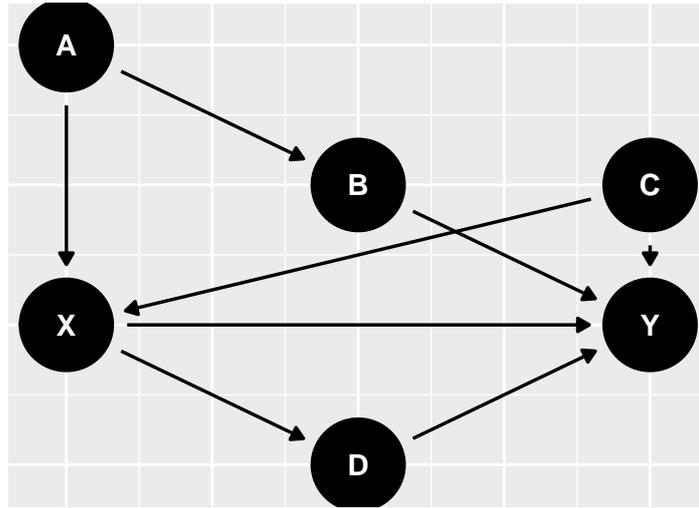


Figura 3.5.: Para medir o efeito causal de X em Y , podemos aplicar o critério backdoor. Neste grafo existem dois caminhos aplicáveis ao critério backdoor: (X, A, B, Y) e (X, C, Y) . No primeiro, A é um confundidor. No segundo caminho, C é um confundidor. Assim, (A, C) bloqueia ambos os caminhos e satisfaz o critério backdoor.

3.2.1. Identificação causal usando o critério backdoor

A seguir, o Teorema 3.19 mostra que, se \mathbf{Z} satisfaz o critério backdoor, então é possível ligar algumas distribuições sob intervenção em X a distribuições observacionais:

Definição 3.18. \mathbf{Z} controla confundidores para medir o efeito causal de X em Y se

$$f(\mathbf{z}|do(x)) = f(\mathbf{z}), \text{ e}$$

$$f(y|do(x), \mathbf{z}) = f(y|x, \mathbf{z}).$$

Teorema 3.19. Se \mathbf{Z} satisfaz o critério backdoor para medir o efeito causal de X em Y , então \mathbf{Z} satisfaz a Definição 3.18.

O Teorema 3.19 mostra que, se \mathbf{Z} satisfaz o critério backdoor, então distribuição de \mathbf{Z} quando aplicamos uma intervenção em X é igual à distribuição marginal de \mathbf{Z} . Além disso, a distribuição condicional de Y dado \mathbf{Z} quando aplicamos uma intervenção em X é igual à distribuição de Y dado X e \mathbf{Z} . Assim, o Teorema 3.19 relaciona distribuições que não geraram os dados a distribuições que os geraram. Com base neste resultado, é possível determinar $f(y|do(x))$ a partir de $f(y, x, \mathbf{z})$:

Corolário 3.20. Se \mathbf{Z} satisfaz a Definição 3.18, então

$$f(y|do(x)) = \int f(y|x, \mathbf{z})f(\mathbf{z})d\mathbf{z}.$$

Para compreender intuitivamente o Corolário 3.20, podemos retornar ao Exemplo 3.5. Considere o caso em que X, Y, Z são as indicadoras de que, respectivamente, o paciente foi submetido ao tratamento, se curou e, é de sexo masculino. Similarmente ao Teorema 3.19, vimos em Exemplo 3.5 que $f(y|do(x))$ é a média de $f(y|x, z)$ ponderada por $f(z)$. Nesta ponderação, utilizamos $f(z)$ ao invés de $f(z|x)$ pois Z é um confundidor e, assim, no modelo

intervencional não propagamos a informação em X por esta variável. A mesma lógica se aplica às variáveis que satisfazem o critério backdoor.

Para calcular quantidades como o ACE (Definição 3.4), utilizamos $\mathbb{E}[Y|do(X)]$. Por meio do Teorema 3.19, é possível obter equivalências entre $\mathbb{E}[Y|do(X)]$ e esperanças obtidas no modelo observacional. Estas equivalências são descritas nos Teoremas 3.21 e 3.22.

Teorema 3.21. *Se \mathbf{Z} satisfaz a Definição 3.18, então*

$$\begin{aligned}\mathbb{E}[g(Y)|do(X = x), \mathbf{Z}] &= \mathbb{E}[g(Y)|X = x, \mathbf{Z}], \text{ e} \\ \mathbb{E}[g(Y)|do(X = x)] &= \mathbb{E}[\mathbb{E}[g(Y)|X = x, \mathbf{Z}]]\end{aligned}$$

Teorema 3.22. *Se \mathbf{Z} satisfaz a Definição 3.18 e X é discreto, então*

$$\begin{aligned}\mathbb{E}[g(Y)|do(X = x), \mathbf{Z}] &= \frac{\mathbb{E}[g(Y)\mathbb{I}(X = x)|\mathbf{Z}]}{f(x|\mathbf{Z})}, \text{ e} \\ \mathbb{E}[g(Y)|do(X = x)] &= \mathbb{E}\left[\frac{g(Y)\mathbb{I}(X = x)}{f(x|\mathbf{Z})}\right]\end{aligned}$$

A seguir, veremos como os Teoremas 3.21 e 3.22 podem ser usados para estimar o efeito causal. Para provar resultados sobre os estimadores obtidos, a seguinte definição será útil

Definição 3.23. Seja \hat{g} um estimador treinado com os dados $(\mathcal{V}_1, \dots, \mathcal{V}_n)$. Dizemos que \hat{g} é invariante a permutações se o estimador não depende da ordem dos dados. Isto é, para qualquer permutações dos índices, $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$, $\hat{g}(\mathcal{V}_1, \dots, \mathcal{V}_n) \equiv \hat{g}(\mathcal{V}_{\pi(1)}, \dots, \mathcal{V}_{\pi(n)})$

Exemplo 3.24. A média amostral é invariante a permutações pois, para qualquer permutação π ,

$$\frac{\sum_{i=1}^n X_i}{n} = \frac{\sum_{i=1}^n X_{\pi(i)}}{n}.$$

3.2.2. Estimação usando o critério backdoor

Fórmula do ajuste

Se \mathbf{Z} satisfaz a Definição 3.18, então $\mathbb{E}[Y|do(X), \mathbf{Z}] = \mathbb{E}[Y|X, \mathbf{Z}]$. Como $\mu(X, Z) := \mathbb{E}[Y|X, \mathbf{Z}]$ é a função de regressão de Y em X e Z , podemos estimar μ utilizando quaisquer métodos de estimação para regressão. Por exemplo, se Y é contínua, possíveis métodos são: regressão linear, Nadaraya-Watson, floresta aleatória de regressão, redes neurais, ... Por outro lado, se Y é discreta, então a função de regressão é estimada por métodos de classificação como: regressão logística, k-NN, floresta aleatória de classificação, redes neurais, ... Para qualquer opção escolhida, denotamos o estimador de μ por $\hat{\mu}$.

Utilizando $\hat{\mu}$, podemos estimar $CACE(\mathbf{Z})$ diretamente. Para tal, note que $CACE(\mathbf{Z})$ é função de $\mathbb{E}[Y|do(X), \mathbf{Z}]$. Como o Teorema 3.21 garante que $\mathbb{E}[Y|do(X = x), \mathbf{Z}] = \mu(x, \mathbf{Z})$, podemos definir o estimador

$$\hat{\mathbb{E}}_1[Y|do(X = x), \mathbf{Z}] := \hat{\mu}(x, \mathbf{Z}).$$

O Teorema 3.21 também orienta a estimação do ACE . Similarmente ao caso anterior, o ACE é função de $\mathbb{E}[Y|do(X)]$. Pelo Teorema 3.21, $\mathbb{E}[Y|do(X = x)] = \mathbb{E}[\mu(x, \mathbf{Z})]$. Assim, se $\hat{\mu} \approx \mu$, $\mathbb{E}[Y|do(X = x)] \approx \mathbb{E}[\hat{\mu}(x, \mathbf{Z})]$.

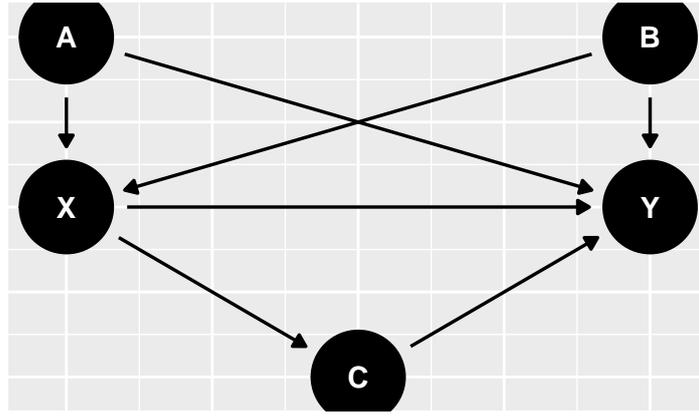


Figura 3.6.: DAG usado como exemplo para estimar efeito de X em Y.

Como $\mathbb{E}[\hat{\mu}(x, \mathbf{Z})]$ é simplesmente uma média populacional, podemos estimá-la com base na média amostral:

$$\hat{\mathbb{E}}_1[Y|do(X = x)] := \frac{\sum_{i=1}^n \hat{\mu}(x, \mathbf{Z}_i)}{n} \approx \mathbb{E}[\hat{\mu}(x, \mathbf{Z})]$$

Definição 3.25. Considere que \mathbf{Z} satisfaz a [Definição 3.18](#) e $\hat{\mu}(x, \mathbf{z})$ é uma estimativa da regressão $\mathbb{E}[Y|X = x, \mathbf{Z} = \mathbf{z}]$. Os estimadores de $\mathbb{E}[Y|do(X = x), \mathbf{Z}]$ e $\mathbb{E}[Y|do(X = x)]$ pela fórmula do ajuste são:

$$\begin{aligned} \hat{\mathbb{E}}_1[Y|do(X = x), \mathbf{Z}] &:= \hat{\mu}(x, \mathbf{Z}) \\ \hat{\mathbb{E}}_1[Y|do(X = x)] &:= \frac{\sum_{i=1}^n \hat{\mu}(x, \mathbf{Z}_i)}{n} \end{aligned}$$

A seguir mostraremos que, se $\hat{\mu}$ converge para μ , então $\hat{\mathbb{E}}_1[Y|do(X = x)]$ converge para $\mathbb{E}[Y|do(X = x)]$. Em outras palavras, é possível utilizar $\hat{\mathbb{E}}_1[Y|do(X = x)]$ para estimar o efeito causal de X em Y por meio de expressões como o *ACE*.

Teorema 3.26. Seja $\mu(X, \mathbf{Z}) := \mathbb{E}[Y|X, \mathbf{Z}]$. Se \mathbf{Z} satisfaz a [Definição 3.18](#), $\mathbb{E}[|\mu(x, \mathbf{Z}_1)|] < \infty$, $\mathbb{E}[|\hat{\mu}(x, \mathbf{Z}_1) - \mu(x, \mathbf{Z}_1)|] = o(1)$, e $\hat{\mu}$ é invariante a permutações ([Definição 3.23](#)), então $\hat{\mathbb{E}}_1[Y|do(X = x)] \xrightarrow{\mathbb{P}} \mathbb{E}[Y|do(X = x)]$.

A seguir, utilizamos dados simulados para ilustrar a implementação da fórmula do ajuste.

Exemplo 3.27. Considere que o grafo causal é dado pela [figur 3.6](#). Vamos supor que os dados são gerados da seguinte forma: $\sigma^2 = 0.01$, $A \sim N(0, \sigma^2)$, $B \sim N(0, \sigma^2)$, $\epsilon \sim \text{Bernoulli}(0.95)$ $X \equiv \mathbb{I}(A + B > 0)\epsilon + \mathbb{I}(A + B < 0)(1 - \epsilon)$, $C \sim N(X, \sigma^2)$, e $Y \sim N(A + B + C + X, \sigma^2)$:

```

# Especificar o grafo
grafo <- dagitty::dagitty("dag {
  X[e] Y[o]
  {A B} -> { X Y }; X -> {C Y}; C -> Y}")

# Simular os dados
n <- 10^5
sd = 0.1

```

```

A <- rnorm(n, 0, sd)
B <- rnorm(n, 0, sd)
eps <- rbinom(n, 1, 0.8)
X <- as.numeric(eps*((A + B) > 0) +
                (1-eps)*((A + B) <= 0))
C <- rnorm(n, X, sd)
Y <- rnorm(n, A + B + C + X, sd)
data <- dplyr::tibble(A, B, C, X, Y)

```

Estimaremos o efeito causal pela fórmula do ajuste (Definição 3.25). Iniciaremos a análise utilizando $\hat{\mu}$ como sendo uma regressão linear simples:

```

# Sejam Z variáveis que satisfazem o critério backdoor para
# estimar o efeito causal de causa em efeito em grafo.
# Retorna uma fórmula do tipo Y ~ X + Z_1 + ... + Z_d
fm_ajuste <- function(grafo, causa, efeito)
{
  var_backdoor <- dagitty::adjustmentSets(grafo)[[1]]
  regressores = c(causa, var_backdoor)
  fm = paste(regressores, collapse = "+")
  fm = paste(c(efeito, fm), collapse = "~")
  as.formula(fm)
}

# Estima E[Efeito/do(causa = x)] pela
# formula do ajuste usando mu_chapeu como regressao
est_do_x_lm <- function(data, mu_chapeu, causa, x)
{
  data %>%
    dplyr::mutate({{causa}} := x) %>%
    predict(mu_chapeu, newdata = .) %>%
    mean()
}

# Estimacão do ACE com regressão linear simples
fm <- fm_ajuste(grafo, "X", "Y")
mu_chapeu_lm <- lm(fm, data = data)
ace_ajuste_lm = est_do_x_lm(data, mu_chapeu_lm, "X", 1) -
  est_do_x_lm(data, mu_chapeu_lm, "X", 0)
round(ace_ajuste_lm)

## [1] 2

```

Em alguns casos, não é razoável supor que $\mathbb{E}[Y|X, \mathbf{Z}]$ é linear. Nestas situações, é fácil adaptar o código anterior

para algum método não-paramétrico arbitrário. Exibimos uma implementação usando XGBoost (?):

```
library(xgboost)
var_backdoor <- dagitty::adjustmentSets(grafo, "X", "Y")[[1]]
mu_chapeu <- xgboost(
  data = data %>%
    dplyr::select(all_of(c(var_backdoor, "X"))) %>%
    as.matrix(),
  label = data %>%
    dplyr::select(Y) %>%
    as.matrix(),
  nrounds = 100,
  objective = "reg:squarederror",
  early_stopping_rounds = 3,
  max_depth = 2,
  eta = .25,
  verbose = FALSE
)

est_do_x_xgb <- function(data, mu_chapeu, causa, x)
{
  data %>%
    dplyr::mutate({{causa}} := x) %>%
    dplyr::select(c(var_backdoor, causa)) %>%
    as.matrix() %>%
    predict(mu_chapeu, newdata = .) %>%
    mean()
}

ace_est_xgb = est_do_x_xgb(data, mu_chapeu, "X", 1) -
  est_do_x_xgb(data, mu_chapeu, "X", 0)
round(ace_est_xgb, 2)

## [1] 2
```

Como o modelo linear era adequado para $\mathbb{E}[Y|X, \mathbf{Z}]$, não vemos diferença entre a estimativa obtida pela regressão linear simples e pelo XGBoost. Mas será que as estimativas estão adequadas? Como simulamos os dados, é possível

calcular diretamente $\mathbb{E}[Y|do(X = x)]$:

$$\begin{aligned}
\mathbb{E}[Y|do(X = x)] &= \mathbb{E}[\mathbb{E}[Y|X = x, A, B]] && \text{Teorema 3.21} \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y|X = x, A, B, C]|X = x, A, B]] && \text{Lei da esperança total} \\
&= \mathbb{E}[\mathbb{E}[A + B + C + X|X = x, A, B]] && Y \sim N(A + B + C + X, \sigma^2) \\
&= \mathbb{E}[A + B + 2x] && C \sim N(X, \sigma^2) \\
&= 2x && \mathbb{E}[A] = \mathbb{E}[B] = 0 \quad (3.5)
\end{aligned}$$

Uma vez calculado $\mathbb{E}[Y|do(X = x)]$, podemos obter o *ACE*:

$$\begin{aligned}
ACE &= \mathbb{E}[Y|do(X = 1)] - \mathbb{E}[Y|do(X = 0)] \\
&= 2 \cdot 1 - 2 \cdot 0 = 2 && \text{linking (3.5)}
\end{aligned}$$

Portanto, as estimativas do *ACE* obtidas pela regressão linear e pelo xgboost estão adequadas.

Ponderação pelo inverso do escore de propensidade (IPW)

Uma outra forma de estimar $\mathbb{E}[Y|do(X = x)]$ e $\mathbb{E}[Y|do(X = x), \mathbf{Z}]$ é motivada pelo [Teorema 3.22](#). Este resultado determina que, se \mathbf{Z} satisfaz a [Definição 3.18](#), então

$$\mathbb{E}[Y|do(X = x), \mathbf{Z}] = \frac{\mathbb{E}[Y\mathbb{I}(X = x)|\mathbf{Z}]}{f(x|\mathbf{Z})}.$$

Na segunda expressão, $\mathbb{E}[Y\mathbb{I}(X = x)|\mathbf{Z}]$ é a regressão de $Y\mathbb{I}(X = x)$ em \mathbf{Z} . Assim, esta quantidade pode ser estimada por um método de regressão arbitrário, que denotaremos por $\widehat{\mathbb{E}}[Y\mathbb{I}(X = x)|\mathbf{Z}]$. Também $f(x|\mathbf{z})$ é usualmente chamado de *escore de propensidade*. Este escore captura a forma como os confundidores atuam sobre X nos dados observacionais. Como f em geral é desconhecido, $f(x|\mathbf{z})$ também o é. Contudo, quando X é discreto, podemos estimar $f(x|\mathbf{z})$ utilizando algum algoritmo arbitrário de classificação. Denotaremos esta estimativa por $\widehat{f}(x|\mathbf{z})$. Se a estimativa for boa, temos

$$\mathbb{E}[Y|do(X = x), \mathbf{Z}] = \frac{\widehat{\mathbb{E}}[Y\mathbb{I}(X = x)|\mathbf{Z}]}{f(x|\mathbf{Z})} \approx \frac{\widehat{\mathbb{E}}[Y\mathbb{I}(X = x)|\mathbf{Z}]}{\widehat{f}(x|\mathbf{z})}.$$

O [Teorema 3.22](#) também orienta a estimação de $\mathbb{E}[Y|do(X = x)]$. Se \mathbf{Z} satisfaz o critério backdoor, então

$$\mathbb{E}[Y|do(X = x)] = \mathbb{E}\left[\frac{Y\mathbb{I}(X = x)}{f(x|\mathbf{Z})}\right]$$

Como nesta expressão a esperança é uma média populacional, ela pode ser aproximada pela média amostral

$$\mathbb{E}\left[\frac{Y\mathbb{I}(X = x)}{f(x|\mathbf{Z})}\right] \approx n^{-1} \sum_{i=1}^n \frac{Y_i\mathbb{I}(X_i = x)}{\widehat{f}(x|\mathbf{Z}_i)}.$$

Combinando estas aproximações, obtemos:

Definição 3.28. Considere que \mathbf{Z} satisfaz a [Definição 3.18](#) e $\widehat{f}(x|\mathbf{z})$ é uma estimativa de $f(x|\mathbf{z})$. Os estimadores

de $\mathbb{E}[Y|do(X = x), \mathbf{Z}]$ e $\mathbb{E}[Y|do(X = x)]$ por IPW são:

$$\widehat{\mathbb{E}}_2[Y|do(X = x), \mathbf{Z}] := \frac{\widehat{\mathbb{E}}[Y\mathbb{I}(X = x)|\mathbf{Z}]}{\widehat{f}(x|\mathbf{z})}$$

$$\widehat{\mathbb{E}}_2[Y|do(X = x)] := n^{-1} \sum_{i=1}^n \frac{Y_i \mathbb{I}(X_i = x)}{\widehat{f}(x|\mathbf{Z}_i)}.$$

Se \widehat{f} converge para f , então sob condições relativamente pouco restritivas $\widehat{\mathbb{E}}_2[Y|do(X = x)]$ converge para $\mathbb{E}[Y|do(X = x)]$.

Teorema 3.29. *Se \mathbf{Z} satisfaz a [Definição 3.18](#), também \widehat{f} é invariante a permutações ([Definição 3.23](#)), $\mathbb{E}[|\widehat{f}(x|\mathbf{Z}_1) - f(x|\mathbf{Z}_1)|] = o(1)$, e existe $M > 0$ tal que $\sup_{\mathbf{z}} \mathbb{E}[|Y|\mathbb{I}(X = x)|\mathbf{Z} = \mathbf{z}] < M$, e existe $\delta > 0$ tal que $\inf_{\mathbf{z}} \min\{f(x|\mathbf{Z}_1), \widehat{f}(x|\mathbf{Z}_1)\} > \delta$, então $\widehat{\mathbb{E}}_2[Y|do(X = x)] \xrightarrow{\mathbb{P}} \mathbb{E}[Y|do(X = x)]$.*

A seguir, utilizamos novamente dados simulados para ilustrar a implementação de IPW:

Exemplo 3.30. Considere que o grafo causal e o modelo de geração dos dados são idênticos àqueles do [Exemplo 3.27](#). Iniciaremos a análise utilizando regressão logística para estimar \widehat{f} .

```
# Sejam Z variáveis que satisfazem o critério backdoor para
# estimar o efeito causal de causa em efeito em grafo.
# Retorna uma fórmula do tipo X ~ Z_1 + ... + Z_d
```

```
fm_ipw <- function(grafo, causa, efeito)
{
  var_backdoor <- dagitty::adjustmentSets(grafo)[[1]]
  fm = paste(var_backdoor, collapse = "+")
  fm = paste(c(causa, fm), collapse = "~")
  as.formula(fm)
}
```

```
# Estimação do ACE por IPW onde
```

```
# Supomos X binário e
```

```
# f_1 é o vetor P(X_i=1|Z_i)
```

```
ACE_ipw <- function(data, causa, efeito, f_1)
```

```
{
  data %>%
  mutate(f_1 = f_1,
         est_1 = {{efeito}}*({causa}==1)/f_1,
         est_0 = {{efeito}}*({causa}==0)/(1-f_1)
  ) %>%
  summarise(do_1 = mean(est_1),
            do_0 = mean(est_0)) %>%
  mutate(ACE = do_1 - do_0) %>%
  dplyr::select(ACE)
}
```

```

fm <- fm_ipw(grafo, "X", "Y")
f_chapeu <- glm(fm, family = "binomial", data = data)
f_1_lm <- predict(f_chapeu, type = "response")
ace_ipw_lm <- data %>% ACE_ipw(X, Y, f_1_lm) %>% as.numeric()
ace_ipw_lm %>% round(2)

## [1] 2.09

```

Também é fácil adaptar o código acima para estimar ACE por IPW utilizando algum método não-paramétrico para estimar \hat{f} . Abaixo há um exemplo utilizando o XGBoost:

```

var_backdoor <- dagitty::adjustmentSets(grafo)[[1]]
f_chapeu <- xgboost(
  data = data %>%
    dplyr::select(all_of(var_backdoor)) %>%
    as.matrix(),
  label = data %>%
    dplyr::select(X) %>%
    as.matrix(),
  nrounds = 100,
  objective = "binary:logistic",
  early_stopping_rounds = 3,
  max_depth = 2,
  eta = .25,
  verbose = FALSE
)

covs <- data %>% dplyr::select(all_of(var_backdoor)) %>% as.matrix()
f_1 <- predict(f_chapeu, newdata = covs)
data %>% ACE_ipw(X, Y, f_1) %>% as.numeric() %>% round(2)

## [1] 1.97

```

Estimador duplamente robusto

Os Teoremas 3.26 e 3.29 mostram que, sob suposições diferentes, $\hat{\mathbb{E}}_1[Y|do(X = x)]$ e $\hat{\mathbb{E}}_2[Y|do(X = x)]$ convergem para $\mathbb{E}[Y|do(X = x)]$. A ideia do estimador duplamente robusto é combinar ambos os estimadores de forma a garantir esta convergência sob suposições mais fracas. Para tal, a ideia por trás do estimador duplamente é que este convirja junto a $\hat{\mathbb{E}}_1[Y|do(X = x)]$ quando este é consistente e para $\hat{\mathbb{E}}_2[Y|do(X = x)]$ quando aquele o é.

Definição 3.31. Se \mathbf{Z} satisfaz a Definição 3.18 e sejam \hat{f} e $\hat{\mu}$ tais quais nas Definições 3.25 e 3.28. O estimador duplamente robusto para $\mathbb{E}[Y|do(X = x)]$, $\hat{\mathbb{E}}_3[Y|do(X = x)]$ é tal que

$$\hat{\mathbb{E}}_3[Y|do(X = x)] = \hat{\mathbb{E}}_1[Y|do(X = x)] + \hat{\mathbb{E}}_2[Y|do(X = x)] - \sum_{i=1}^n \frac{\mathbb{I}(X_i = x)\hat{\mu}(x, \mathbf{Z}_i)}{n\hat{f}(x|\mathbf{Z}_i)}$$

O estimador duplamente robusto é consistente para $\mathbb{E}[Y|do(X = x)]$ tanto sob as condições do [Teorema 3.26](#) quanto sob as do [Teorema 3.29](#). A ideia básica é que, sob as condições do [Teorema 3.26](#), $\widehat{\mathbb{E}}_1[Y|do(X = x)]$ é consistente para $\mathbb{E}[Y|do(X = x)]$ e $\widehat{\mathbb{E}}_2[Y|do(X = x)] - \sum_{i=1}^n \frac{\mathbb{I}(X_i=x)\widehat{\mu}(x, \mathbf{Z}_i)}{n\widehat{f}(x|\mathbf{Z}_i)}$ converge para 0. Isto é, quando $\widehat{\mathbb{E}}_1[Y|do(X = x)]$ é consistente, o estimador duplamente robusto seleciona este termo. Similarmente, sob as condições do [Teorema 3.29](#), $\widehat{\mathbb{E}}_2[Y|do(X = x)]$ é consistente para $\mathbb{E}[Y|do(X = x)]$ e $\widehat{\mathbb{E}}_1[Y|do(X = x)] - \sum_{i=1}^n \frac{\mathbb{I}(X_i=x)\widehat{\mu}(x, \mathbf{Z}_i)}{n\widehat{f}(x|\mathbf{Z}_i)}$ converge para 0.

Teorema 3.32. *Suponha que existe $\epsilon > 0$ tal que $\inf_{\mathbf{z}} \widehat{f}(x|\mathbf{z}) > \epsilon$, existe $M > 0$ tal que $\sup_{\mathbf{z}} \widehat{\mu}(x, \mathbf{z}) < M$, e $\widehat{\mu}$ e \widehat{f} são invariantes a permutações ([Definição 3.23](#)). Se as condições do [Teorema 3.26](#) ou do [Teorema 3.29](#) estão satisfeitas, então*

$$\widehat{\mathbb{E}}_3[Y|do(X = x)] \xrightarrow{\mathbb{P}} \mathbb{E}[Y|do(X = x)].$$

Exemplo 3.33 (Estimador duplamente robusto). Considere que o grafo causal e o modelo de geração dos dados são iguais àqueles descritos no [Exemplo 3.27](#). Para implementar o estimador duplamente robusto combinaremos o estimador da fórmula do ajuste obtido por regressão linear no [Exemplo 3.27](#) e aquele de IPW por regressão logística no [Exemplo 3.30](#).

```
mu_1_lm <- data %>%
  dplyr::mutate(X = 1) %>%
  predict(mu_chapeu_lm, newdata = .)
mu_0_lm <- data %>%
  dplyr::mutate(X = 0) %>%
  predict(mu_chapeu_lm, newdata = .)
corr <- data %>%
  mutate(mu_1 = mu_1_lm,
         mu_0 = mu_0_lm,
         f_1 = f_1_lm,
         corr_1 = (X == 1)*mu_1/f_1,
         corr_0 = (X == 0)*mu_0/(1-f_1)) %>%
  summarise(corr_1 = mean(corr_1),
            corr_0 = mean(corr_0)) %>%
  mutate(corr = corr_1 - corr_0) %>%
  dplyr::select(corr) %>%
  as.numeric()
ace_rob_lm <- ace_ajuste_lm + ace_ipw_lm - corr
ace_rob_lm %>% round(2)

## [1] 2
```

3.2.3. Exercícios

Exercício 3.34. Prove o [Lema 3.16](#).

Exercício 3.35. Prove o [Lema 3.17](#).

Exercício 3.36. Prove que se $X \notin Anc(Y)$, então $ACE = 0$.

Exercício 3.37. Prove que a variância amostral satisfaz o [Definição 3.23](#).

Exercício 3.38. Utilizando como referência o grafo e o código no [Exemplo 3.27](#), simule dados tais que a estimativa do ACE é diferente quando um método de regressão linear e um de regressão não-paramétrica são usados.

3.2.4. Regression Discontinuity Design (RDD)

Em determinadas situações, X é completamente determinado pelos confundidores, \mathbf{Z} (?). Por exemplo, considere que desejamos determinar o efeito causal que um determinado programa social do governo traz sobre o nível de educação dos cidadãos. Neste caso, X é a indicadora de que o indivíduo é elegível ao programa e Y mede o seu nível de educação. Em alguns casos, é razoável supor que X é completamente determinado por Z , a renda do indivíduo.

A situação acima traz desafios para a fórmula do ajuste e IPW discutidos anteriormente. Primeiramente, como $X = h(\mathbf{Z})$, não é possível estimar $\mathbb{E}[Y|X = x, \mathbf{Z} = \mathbf{z}]$ quando $x \neq h(\mathbf{z})$. Portanto, não é possível utilizar a fórmula do ajuste, uma vez que ela se baseia na expressão $\mathbb{E}[\mathbb{E}[Y|X = x, \mathbf{Z}]]$. Similarmente, o estimador de IPW envolve uma divisão por $f(x|\mathbf{Z})$. Assim, quando $x \neq h(\mathbf{Z})$ há uma divisão por 0, o que torna o estimador indefinido.

Identificação causal no RDD

Apesar destas dificuldades, é possível medir nestas situações parte do efeito causal de X em Y . Suponha que $\mathbf{Z} \in \mathfrak{R}$ e que existe z_1 tal que $X = \mathbb{I}(\mathbf{Z} \geq \mathbf{z}_1)$. Por exemplo, um benefício pode estar disponível apenas para cidadãos que tenham renda abaixo de um teto ou uma lei pode ter efeitos a partir de uma determinada data.

Neste caso, podemos estar interessados em $\mathbb{E}[Y|do(X = x), \mathbf{Z} = \mathbf{z}_1]$, o efeito causal que X tem na fronteira de sua implementação. Intuitivamente, próximo a esta fronteira, as unidades amostrais são todas similares em relação aos confundidores. Assim, se na fronteira houver uma diferença em Y entre os valores de X , esta diferença deve ser decorrente do efeito causal de X . Esta intuição é formalizada no resultado de identificação causal abaixo:

Teorema 3.39 (?). Considere que $\mathbf{Z} \in \mathfrak{R}$ satisfaz a [Definição 3.18](#), $X \in \{0, 1\}$, e $\mathbb{E}[Y|do(X = 0), \mathbf{Z}]$ e $\mathbb{E}[Y|do(X = 1), \mathbf{Z}]$ são contínuas em $\mathbf{Z} = \mathbf{z}_1$.

Se $X \equiv \mathbb{I}(\mathbf{Z} \geq \mathbf{z}_1)$, então

$$CACE(\mathbf{Z} = \mathbf{z}_1) = \lim_{\mathbf{z} \downarrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}].$$

Se $f(x|\mathbf{Z}) \in (0, 1)$ é contínua exceto em \mathbf{z}_1 , então

$$CACE(\mathbf{Z} = \mathbf{z}_1) = \frac{\lim_{\mathbf{z} \downarrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}]}{\lim_{\mathbf{z} \downarrow \mathbf{z}_1} f(X = 1|\mathbf{Z} = \mathbf{z}) - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} f(X = 1|\mathbf{Z} = \mathbf{z})}.$$

Um detalhe sutil do [Teorema 3.39](#) é que $X \equiv \mathbb{I}(\mathbf{Z} > \mathbf{z}_1)$ não é o suficiente para termos certeza que \mathbf{Z} satisfaz o critério backdoor. Por exemplo, considere que o governo criasse um benefício fiscal para todas empresas sediadas em um determinado município. Neste caso, a ocorrência do benefício é função da sede da empresa. Contudo, a relação causal é mais complexa. Se o benefício for suficientemente alto, poderia motivar empresas a moverem sua sede para o município. Em outras palavras, o benefício seria causa da localização da sede e não o contrário. Neste caso, não seria possível aplicar o [Teorema 3.39](#). Este tipo de raciocínio indica que a análise por RDD é mais efetiva quando é difícil interferir sobre o valor de \mathbf{Z} . Por exemplo, como um indivíduo não pode interferir sobre a

sua idade, é mais fácil justificar o uso de RDD em uma campanha de vacinação em que apenas indivíduos acima de uma determinada idade são vacinados.

Um outro ponto importante de interpretação do [Teorema 3.39](#) é que, embora $\mathbb{E}[Y|do(X = 0), \mathbf{Z}]$ e $\mathbb{E}[Y|do(X = 1), \mathbf{Z}]$ sejam supostas contínuas, $f(X = 1|\mathbf{Z})$ e $\mathbb{E}[Y|\mathbf{Z}]$ não o são. Intuitivamente, podemos imaginar X representa a indicadora de que uma determinada política é adotada. Por exemplo, podemos imaginar que X indica que um indivíduo foi vacinado, \mathbf{Z} a sua idade e Y a sua hospitalização. Neste caso, $\mathbb{E}[Y|do(X = 0), \mathbf{Z}]$ e $\mathbb{E}[Y|do(X = 1), \mathbf{Z}]$ representam a taxa de hospitalização quando todos os indivíduos são vacinados ou quando todos eles não o são. Nestas situações, seria razoável supor que a taxa de hospitalização é contínua em função da idade, pois não esperamos que exista uma grande descontinuidade nas condições de saúde entre indivíduos com 69 e com 70 anos de idade. Este tipo de conclusão muitas vezes é resumido pela expressão em latim *natura non facit saltus* (a natureza não faz saltos). Por outro lado, nos dados observados, a política não é adotada para uma faixa de valores de \mathbf{Z} e passa a ser adotada a partir de um ponto, o que é responsável pela descontinuidade em $\mathbb{E}[Y|\mathbf{Z}]$ e em $f(X = 1|\mathbf{Z})$. Podemos imaginar que a vacinação é empregada somente em indivíduos com mais de 70 anos. Esta descontinuidade na política humana cria uma diferença importante entre indivíduos com 69 e com 70 anos, o que explica uma diferença grande nas taxas de hospitalização entre estas idades nos dados observados.

Estimação no RDD

O [Teorema 3.39](#) indica que $CACE(\mathbf{Z} = \mathbf{z}_1)$ é função da regressão de Y sobre \mathbf{Z} , $\mathbb{E}[Y|\mathbf{Z}]$, e sobre o classificador, $f(X = 1|\mathbf{Z})$. Uma possível estratégia é estimarmos estas quantidades separadamente e, a seguir, estimarmos o $CACE$ trocando as quantias populacionais pelas quantias estimadas.

Uma dificuldade nesta estratégia é que sabemos que $\mathbb{E}[Y|\mathbf{Z}]$ e $f(X = 1|\mathbf{Z})$ são descontínuas. Para lidar com esta dificuldade, uma possibilidade é realizar uma regressão para $\mathbf{Z} < \mathbf{z}_1$ e outra para $\mathbf{Z} \geq \mathbf{z}_1$.

Definição 3.40. Seja $D_< = \{i : \mathbf{Z}_i < \mathbf{z}_1\}$ o conjunto de unidades amostrais em que $\mathbf{Z}_i < \mathbf{z}_1$, $\widehat{\mathbb{E}}_<[Y|\mathbf{Z}]$ e $\widehat{f}_<(X = 1|\mathbf{Z})$ regressões ajustadas utilizando apenas dados em $D_<$ e $\widehat{\mathbb{E}}_\geq[Y|\mathbf{Z}]$ e $\widehat{f}_\geq(X = 1|\mathbf{Z})$ ajustadas em D_\geq^c . O estimador RDD para $CACE(\mathbf{z}_1)$ é

$$\widehat{CACE}(\mathbf{z}_1) := \frac{\widehat{\mathbb{E}}_\geq[Y|\mathbf{z}_1] - \widehat{\mathbb{E}}_<[Y|\mathbf{z}_1]}{\widehat{f}_\geq(X = 1|\mathbf{Z}) - \widehat{f}_<(X = 1|\mathbf{Z})}.$$

Em particular, se sabemos a priori que $f(X = 1|\mathbf{z}) = 1$ para $\mathbf{z} \geq \mathbf{z}_1$ e $f(X = 1|\mathbf{z}) = 0$ para $\mathbf{z} < \mathbf{z}_1$, então

$$\widehat{CACE}(\mathbf{z}_1) := \widehat{\mathbb{E}}_\geq[Y|\mathbf{z}_1] - \widehat{\mathbb{E}}_<[Y|\mathbf{z}_1]$$

O exemplo a seguir ilustra a implementação de RDD quando $X \equiv \mathbb{I}(\mathbf{Z} \geq \mathbf{z}_1)$ utilizando tanto regressão linear quanto regressão de Kernel de Nadaraya-Watson.

Exemplo 3.41. Considere que Z_i satisfaz o critério backdoor para estimar o efeito causal de X em Y . Além disso, $Z_i \sim N(0, 1)$, $X_i \equiv \mathbb{I}(Z_i \geq 0)$ e $Y_i|X_i, Z_i \sim N(50(X_i + 1)(Z_i + 1), 1)$. Podemos simular os dados da seguinte forma:

```
n <- 1000
Z <- rnorm(n)
X <- Z >= 0
Y <- rnorm(n, 50*(X+1)*(Z+1))
```

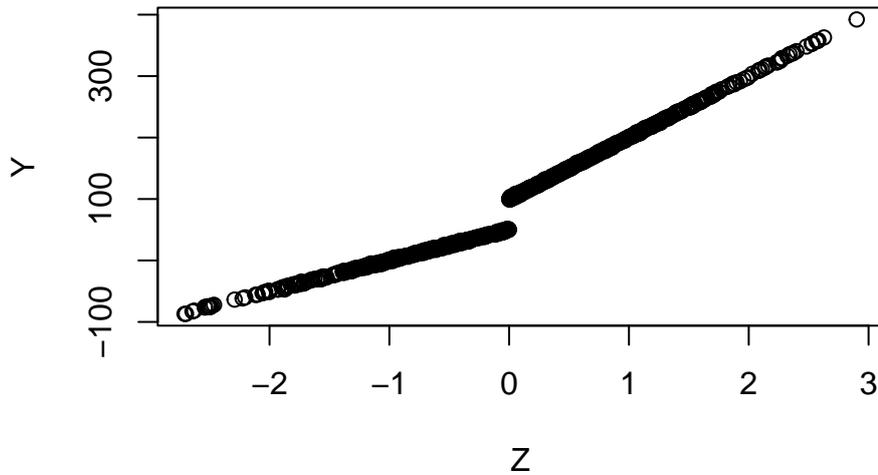


Figura 3.7.: Exemplo em que Z satisfaz o critério backdoor para medir o efeito causal de X em Y e $X = I(Z > 0)$. Como resultado da descontinuidade da propensidade de X em $Z = 0$, há uma descontinuidade na regressão de Y em Z no ponto $Z=0$.

```
data <- tibble(X, Y, Z)
plot(Z, Y)
```

Como estamos simulando os dados, podemos calcular $CACE(0)$:

$$\begin{aligned}
 CACE(0) &= \lim_{z \downarrow 0} \mathbb{E}[Y | \mathbf{Z} = \mathbf{z}] - \lim_{z \uparrow 0} \mathbb{E}[Y | \mathbf{Z} = \mathbf{z}] \\
 &= \lim_{z \downarrow 0} \mathbb{E}[Y | X = 1, \mathbf{Z} = \mathbf{z}] - \lim_{z \uparrow 0} \mathbb{E}[Y | X = 0, \mathbf{Z} = \mathbf{z}] \\
 &= \lim_{z \downarrow 0} 50(1 + 1)(z + 1) - \lim_{z \uparrow 0} 50(0 + 1)(z + 1) = 50
 \end{aligned}$$

Teorema 3.39

O código abaixo estima $CACE(0)$ usando regressão linear:

```
regs = data %>%
  mutate(Z1 = (Z >= 0)) %>%
  group_by(Z1) %>%
  summarise(
    intercepto = lm(Y ~ Z)$coefficients[1],
    coef_angular = lm(Y ~ Z)$coefficients[2]
  )
regs

## # A tibble: 2 x 3
##   Z1     intercepto coef_angular
##   <lg1>      <dbl>      <dbl>
```

```
## 1 FALSE      50.1      50.1
## 2 TRUE       99.9     100.

est_cace = 1*regs[2, 2] + 0*regs[2, 3] -
  1*regs[1, 2] + 0*regs[1, 3]
round(as.numeric(est_cace), 2)

## [1] 49.81
```

Similarmente, podemos estimar $CACE(0)$ usando regressão por kernel de Nadaraya-Watson:

```
library(np)
options(np.messages = FALSE)
nw_reg <- function(data, valor)
{
  bw <- npregbw(xdat = data$Z, ydat = data$Y)$bw
  npksum(txdat= data$Z, exdat = valor, tydat = data$Y, bws = bw)$ksum/
  npksum(txdat = data$Z, exdat = valor, bws = bw)$ksum
}

reg_baixo <- data %>%
  filter(Z < 0) %>%
  nw_reg(0)
reg_cima <- data %>%
  filter(Z >= 0) %>%
  nw_reg(0)
est_cace <- reg_cima - reg_baixo
round(est_cace, 2)

## [1] 51.31
```

3.2.5. Exercícios

Exercício 3.42. Crie um exemplo em que, ao contrário do [Exemplo 3.41](#), $\mathbb{E}[Y|X = 1, \mathbf{Z}]$ não é linear em \mathbf{Z} . Compare as estimativas de $CACE$ usando a regressão linear e algum método de regressão não-paramétrica.

3.3. Controlando mediadores (critério frontdoor)

Há casos em que não existem variáveis observadas que satisfazem o critério backdoor. Por exemplo, considere o grafo causal na [figur 3.8](#) (?). Neste grafo, estamos interessados em compreender o efeito causal do fumo (X) sobre a incidência de câncer (Y). Além disso, fatores genéticos não observáveis (G) são um potencial confundidor, uma vez que podem ter influência tanto sobre o fumo quanto sobre a incidência de câncer. Assim, como G não é observado, não é possível implementar os métodos de estimação vistos na última seção. Apesar desta dificuldade, ainda é possível medir o efeito causal de X em Y na [figur 3.8](#).

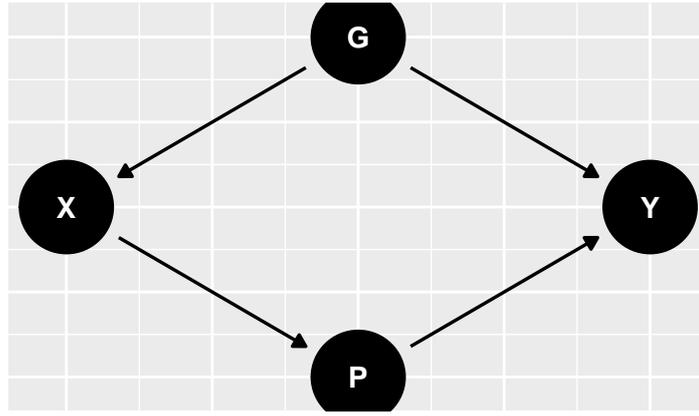


Figura 3.8.: .

Para tal, primeiramente observe que é possível estimar o efeito causal de X em P e de P em Y . Para medir o efeito causal de X em P , note que \emptyset satisfaz o critério backdoor. Isso ocorre pois Y é um colisor em $X \leftarrow G \rightarrow Y \leftarrow P$. Além disso, como $X = Pa(P)$, decorre do [Lema 3.17](#) que X satisfaz o critério backdoor para medir o efeito causal de P em Y . Das duas últimas conclusões decorre do [Teorema 3.19](#) que $f(P|do(X)) = f(P|X)$ e que $f(Y|do(P)) = \int f(Y|P, X)f(X)dX$.

A seguir, o critério frontdoor consiste em observar que P está no único caminho direcionado de X a Y , $X \rightarrow P \rightarrow Y$. Assim, é possível provar a identificação causal

$$\begin{aligned} f(Y|do(X)) &= \int f(P|do(X))f(Y|do(P))dP \\ &= \int f(P|do(X)) \int f(Y|P, X)f(X)dX. \end{aligned}$$

O critério frontdoor é formalizado a seguir:

Definição 3.43. \mathbf{W} satisfaz o critério frontdoor para medir o efeito causal de X em Y se:

1. para todo caminho direcionado de X em Y , C , existe $C_i \in \mathbf{W}$ e, para todo $W \in \mathbf{W}$, existe caminho direcionado de X em Y , C , e i tal que $C_i = W$.
2. \emptyset satisfaz o item 2 do critério backdoor ([Definição 3.11](#)) para medir o efeito causal de X em \mathbf{W} .
3. X satisfaz o item 2 do critério backdoor ([Definição 3.11](#)) para medir o efeito causal de \mathbf{W} em Y .

A [Definição 3.43](#) elenca todos os itens que utilizamos na análise da [figur 3.8](#). O primeiro item do critério identifica que \mathbf{W} deve interceptar todos os caminhos direcionados de X a Y . Isto é, \mathbf{W} capturar a informação de todos os mediadores de X a Y . O segundo e terceiro itens estabelecem as condições para que seja possível aplicar o critério backdoor para identificar $f(\mathbf{W}|do(X))$ e $f(Y|do(\mathbf{W}))$.

Identificação causal

O critério frontdoor possibilita a identificação do efeito causal de X em Y :

Teorema 3.44. Se \mathbf{W} satisfaz o critério frontdoor para medir o efeito causal de X em Y , então

$$f(Y|do(X = x)) = \int f(\mathbf{W}|x) \int f(Y|\mathbf{W}, X) f(X) dX d\mathbf{W}$$

Teorema 3.45. Se \mathbf{W} satisfaz o critério frontdoor para estimar o efeito causal de X em Y , então

$$\mathbb{E}[Y|do(X = x)] = \mathbb{E} \left[\frac{Y \cdot f(W|x)}{f(W|X)} \right]$$

Estimação pelo critério frontdoor

A estimação é um tema menos desenvolvido ao aplicar o critério frontdoor. Alguns estimadores não-paramétricos são apresentados em ?. A seguir, desenvolvemos um estimador não-paramétrico mais simples inspirado na estratégia de IPW.

Definição 3.46. Considere que \mathbf{W} satisfaz o critério frontdoor para medir o efeito causal de X em Y e que $\widehat{f}(\mathbf{W}|X)$ é um estimador de $f(\mathbf{W}|X)$. Um estimador do tipo IPW para $\mathbb{E}[Y|do(X = x)]$ é dado por

$$\widehat{\mathbb{E}}_f[Y|do(X = x)] := n^{-1} \sum_{i=1}^n \frac{Y_i \widehat{f}(\mathbf{W}_i|x)}{\widehat{f}(\mathbf{W}_i|X_i)}.$$

Para provar o [Teorema 3.44](#) utilizamos o *do calculus*, que é discutido na [Seção 3.4](#).

3.4. Do-calculus

O *do calculus* consiste em um conjunto de regras para alterar densidade envolvendo o operador “do”. Por exemplo, o *do calculus* explica como remover o operador do, trocá-lo pelo condicionamento simples, ou remover algum condicionamento simples. Para apresentar o *do calculus*, é necessário primeiramente definir algumas modificações sobre o grafo causal.

Definição 3.47. Seja (\mathcal{G}, f) um CM tal que $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

$$\begin{aligned} \mathcal{G}(\bar{\mathbb{V}}) &:= (\mathcal{V}, \{E \in \mathcal{E} : E_2 \notin \mathbb{V}\}) \\ \mathcal{G}(\bar{\mathbb{V}}_1, \underline{\mathbb{V}}_2) &:= (\mathcal{V}, \{E \in \mathcal{E} : E_2 \notin \mathbb{V}_1 \text{ e } E_1 \notin \mathbb{V}_2\}) \\ \mathcal{G}(\bar{\mathbb{V}}_1, \mathbb{V}_2^+) &:= (\mathcal{V} \cup \{I_V : V \in \mathbb{V}_2\}, \{E \in \mathcal{E} : E_2 \notin \mathbb{V}_1\} \cup \{(I_V, V) : V \in \mathbb{V}_2\}) \end{aligned}$$

Isto é, $\mathcal{G}(\bar{\mathbb{V}})$ é o grafo obtido retirando de \mathcal{G} as arestas que apontam para \mathbb{V} , $\mathcal{G}(\bar{\mathbb{V}}_1, \underline{\mathbb{V}}_2)$ é o grafo obtido retirando de \mathcal{G} todas as arestas que apontam para \mathbb{V}_1 ou que saem de \mathbb{V}_2 , e $\mathcal{G}(\bar{\mathbb{V}}_1, \mathbb{V}_2^+)$ é o grafo obtido adicionando a \mathcal{G} um novo vértice I_V e uma aresta $I_V \rightarrow V$, para todo $V \in \mathbb{V}_2$, e retirando todas as arestas que apontam para \mathbb{V}_1 .

Com base na [Definição 3.47](#), é possível apresentar o *do calculus*:

Teorema 3.48. Seja (\mathcal{G}, f) um CM e \mathbf{X} , \mathbf{Y} , \mathbf{W} e \mathbf{Z} conjuntos de vértices disjuntos:

1. Se $\mathbf{Y} \perp^d \mathbf{Z} | \mathbf{X} \cup \mathbf{W}$ em $\mathcal{G}(\bar{\mathbf{X}})$, então $f(\mathbf{Y}|do(\mathbf{X}), \mathbf{Z}, \mathbf{W}) = f(\mathbf{Y}|do(\mathbf{X}), \mathbf{W})$.
2. Se $\mathbf{Y} \perp^d \mathbf{W} | \mathbf{Z} \cup \mathbf{X}$ em $\mathcal{G}(\bar{\mathbf{X}}, \underline{\mathbf{W}})$, então $f(\mathbf{Y}|do(\mathbf{X}), do(\mathbf{W}), \mathbf{Z}) = f(\mathbf{Y}|do(\mathbf{X}), \mathbf{W}, \mathbf{Z})$.
3. Se $\mathbf{Y} \perp^d I_{\mathbf{X}} | \mathbf{Z} \cup \mathbf{W}$ em $\mathcal{G}(\bar{\mathbf{W}}, \mathbf{X}^+)$, então $f(Y|do(\mathbf{W}), do(\mathbf{X}), \mathbf{Z}) = f(Y|do(\mathbf{W}), \mathbf{Z})$.

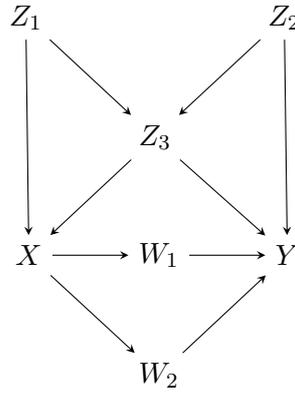


Figura 3.9.: Modelo estrutural causal do [Exercício 3.50](#)

O seguinte lema mostra como o do calculus generaliza certos aspectos do critério backdoor:

Lema 3.49. *X* satisfaz o item 2 do critério backdoor para medir o efeito causal de \mathbf{W} em Y se e somente se $Y \perp^d \mathbf{W} | X$ em $\mathcal{G}(\mathbf{W})$.

Utilizando o do calculus, é possível obter todas as relações de identificação que são válidas supondo apenas que f é compatível com o grafo causal (??). Contudo, às vezes é razoável fazer mais suposições. Discutiremos este tipo de situação no próximo capítulo.

3.4.1. Exercícios

Exercício 3.50 ([p.48]). Considere o modelo estrutural causal em [figur 3.9](#).

- Para cada um dos pares de variáveis a seguir, determine um conjunto de outras variáveis que as d-separa: (Z_1, W) , (Z_1, Z_2) , (Z_1, Y) , (Z_3, W) , e (X, Y) .
- Para cada par de variáveis no item anterior, determine se elas são d-separadas dado todas as demais variáveis.
- Determine conjuntos de variáveis que satisfazem, respectivamente, o critério backdoor e o critério frontdoor para estimar o efeito causal de X em Y .
- Considere que para cada variável, V , temos que $V \equiv \beta_V \cdot Pa(V) + \epsilon_V$, onde os ϵ são i.i.d. e normais padrão e β_V são vetores conhecidos. Isto é, a distribuição de cada variável é determinada através de uma regressão linear simples em seus pais. Determine $f(Y|do(X = x))$ utilizando a fórmula do ajuste nos 2 casos abordados no item anterior.

Exercício 3.51. Considere que $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ é um grafo causal e $\mathbf{X}, \mathbf{W}, \mathbf{Y} \subseteq \mathcal{V}$. Além disso, para todo caminho, $C = (C_1, \dots, C_n)$, com $C_1 = X \in \mathbf{X}$, $C_n = Y \in \mathbf{Y}$, e com $X \rightarrow C_2$, C está bloqueado dado \mathbf{W} . Prove que $f(\mathbf{y}|do(\mathbf{X})) = \int f(\mathbf{y}|\mathbf{w})f(\mathbf{w}|do(\mathbf{X}))d\mathbf{w}$ e $\mathbb{E}[Y|do(\mathbf{X})] = \mathbb{E}[\mathbb{E}[Y|\mathbf{W}]|do(\mathbf{X})]$.

Exercício 3.52. Prove que se \mathbf{W} satisfaz o critério frontdoor para medir o efeito causal de X em Y , então $f(\mathbf{W}|do(X)) = f(\mathbf{W}|X)$ e $f(Y|do(\mathbf{W})) = \int f(Y|\mathbf{W}, X = x^*)f(X = x^*)dx^*$.

Exercício 3.53. Prove o [Lema 3.49](#).

4. Resultados potenciais

No capítulo passado, vimos que $f(y|do(x))$ nos permite entender o comportamento de Y em um cenário distinto dos dados observados. Por exemplo, se X é a indicadora de um tratamento e Y é a indicadora de cura, então $f(y|do(X = 1))$ nos permite entender a proporção de cura em um cenário hipotético em que administramos o tratamento a todos os indivíduos. Esta distribuição nos permite investigar questões causais que não eram acessíveis usando apenas a distribuição observacional, $f(y, x)$.

Contudo, algumas perguntas causais não são respondidas utilizando apenas os mecanismos desenvolvidos no [capítulo 3](#). Por exemplo, qual a probabilidade de que um indivíduo se cure quando recebe o tratamento e não se cure quando não o recebe. Quando tentamos traduzir esta questão, notamos que partes dela envolvem $Y = 1$ e $do(X = 1)$ e outras partes envolvem $Y = 0$ e $do(X = 0)$. Se tentarmos uma tradução ingênua, podemos obter uma expressão como $\mathbb{P}(Y = 1, Y = 0 | do(X = 1), do(X = 0))$. Contudo, a probabilidade acima não responde à pergunta colocada. Em primeiro lugar, não está definido fazermos as intervenções $do(X = 1)$ e $do(X = 0)$ na mesma unidade amostral. Além disso, mesmo que a probabilidade estivesse definida, é impossível que o mesmo Y assumo tanto o valor 1 quanto 0. Isto é, $\mathbb{P}(Y = 1, Y = 0 | \dots) = 0$.

A última constatação nos revela que o modelo no [capítulo 3](#) não tem variáveis suficientes para traduzir a pergunta levantada. Se imaginamos que é possível que um indivíduo se cure ao receber o tratamento e não se cure quando não o recebe, isto ocorre pois as ocorrências de cura em cada cenário hipotético não são logicamente equivalentes. Em outras palavras, é como se houvessem *resultados potenciais*¹, Y_1 e Y_0 , para indicar a ocorrência de cura em cada cenário considerado. Com o uso destas variáveis, poderíamos escrever $\mathbb{P}(Y_1 = 1, Y_0 = 0)$.

O objetivo desta seção é incluir este tipo de variável de forma a preservar as ferramentas desenvolvidas no [capítulo 3](#).² Neste quesito, a maior dificuldade será estabelecer a distribuição conjunta entre os resultados potenciais. Para tal, será útil relembrar um lema fundamental em simulação:

Lema 4.1. *Considere que $F(v|Pa(V))$ é uma função de densidade acumulada condicional arbitrária e $U \sim U(0, 1)$. Se definirmos, $V \equiv F^{-1}(U|Pa(V))$, então $V|Pa(V) \sim F$.*

O [Lema 4.1](#) traz várias interpretações que nos serão úteis. A primeira interpretação, de caráter técnico, é que podemos simular de qualquer distribuição multivariada utilizando apenas variáveis i.i.d. e funções determinísticas. Em particular, podemos reescrever um SCM de tal forma que cada vértice, V , seja função determinística de seus pais e uma variável de ruído, U_V . Esta abordagem, que está ligada a modelos de equações estruturais, é apresentada nas [Definições 4.2](#) e [4.3](#).

Definição 4.2. Seja $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ um grafo causal. O grafo causal estrutural, $\mathcal{G}^+ = (\mathcal{V}^+, \mathcal{E}^+)$, é tal que $\mathcal{V}^+ = \mathcal{V} \cup (U_V)_{V \in \mathcal{V}}$ e $\mathcal{E}^+ = \mathcal{E} \cup \{(U_V, V) : V \in \mathcal{V}\}$. Isto é, para cada $V \in \mathcal{V}$, \mathcal{G}^+ adiciona uma nova variável U_V e uma aresta de U_V a V .

¹Esta é uma tradução livre da expressão “potential outcomes” usada em inglês.

²Para tal, adotaremos uma construção baseada em ?.

Definição 4.3. Seja (\mathcal{G}, f) um CM. O Modelo Estrutural Causal (SCM) para (\mathcal{G}, f) , (\mathcal{G}^+, f^+) , é tal que \mathcal{G}^+ é o grafo causal estrutural de \mathcal{G} , $(U_V)_{V \in \mathcal{V}}$ são independentes segundo f^+ e, para cada $V \in \mathcal{V}$, existe uma função determinística, $g_V : U_V \times Pa(V) \rightarrow \mathfrak{R}$, tal que $f^+(V|U_V, Pa(V)) = \mathbb{I}(V = g_V(U_V, Pa(V)))$ e $f^+(\mathcal{V}) = f(\mathcal{V})$.

O [Exemplo 4.4](#) ilustra uma forma de obter um SCM em equações estruturais a partir de um SCM com dois vértices.

Exemplo 4.4. Considere que $X \rightarrow Y$, $X \sim \text{Exp}(1)$ e $Y|X \sim \text{Exp}(X)$. Neste caso, o grafo estrutural causal é dado por $U_X \rightarrow X \rightarrow Y \leftarrow U_Y$. Além disso, existem várias representações do SCM em equações estruturais. Uma possibilidade é definir que U_X e U_Y são i.i.d. e $U(0, 1)$, $X \equiv -\log(U_X)$ e $Y \equiv -\log(U_Y)/X$.

O [Lema 4.1](#) também permite uma interpretação de caráter mais filosófico. Podemos imaginar que toda variável em um SCM, V , é uma função determinística de seus pais e de *condições locais* não-observadas, U_V . A expressão “condições locais” indica que cada U_V é usada somente para gerar V e que as variáveis em U são independentes, isto é, não trazem informação umas sobre as outras.

A interpretação acima é usada na definição de resultados potenciais. A ideia principal é que as mesmas funções determinísticas e variáveis de ruído locais são usadas para gerar todos os resultados potenciais. A única diferença é que, para cada resultado potencial, o valor das variáveis em que houve intervenção é fixado. Esta definição é compatível com a ideia de que não é possível modificar os ruídos locais por meio da intervenção. Em outras palavras, o resultado potencial é o mais próximo possível do resultado observado sob a restrição que fixamos os valores das variáveis em que houve intervenção.

Definição 4.5. Seja (\mathcal{G}, f) um CM de grafo causal $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ e (\mathcal{G}^+, f^+) o seu SCM. O grafo de resultados potenciais dado por intervenções em $\mathbf{X} \subseteq \mathcal{V}$, $\mathcal{G}^* = (\mathcal{V}^*, \mathcal{E}^*)$ é tal que

$$\begin{aligned}\mathcal{V}^* &= \{W_{\mathbb{V}=\mathbf{v}} : W \in \mathcal{V}, \mathbb{V} \subseteq \mathbf{X}, \mathbf{v} \in \text{supp}(\mathbb{V})\} \cup \{U_W : W \in \mathcal{V}\}, \\ \mathcal{E}^* &= \{(W_{\mathbb{V}=\mathbf{v}}, Z_{\mathbb{V}=\mathbf{v}}) : \mathbb{V} \subseteq \mathbf{X}, \mathbf{v} \in \text{supp}(\mathbb{V}), (W, Z) \in \mathcal{E}^+, Z \notin \mathbb{V}\}.\end{aligned}$$

Para todo $W \in \mathcal{V}$, abreviamos W_\emptyset por W .

Em palavras, o grafo de resultados potenciais cria uma cópia de \mathcal{G} para cada possível intervenção, $\mathbb{V} = \mathbf{v}$. Além disso, adiciona-se uma aresta de U_W para cada cópia de W . Esta construção indica que as mesmas variáveis em U geram todos os resultados potenciais. Também, para cada vértice em que houve uma intervenção, $W_{\mathbb{V}=\mathbf{v}} \in \mathbb{V}_{\mathbf{v}=\mathbf{v}}$, removem-se todas as arestas que apontam para $W_{\mathbb{V}=\mathbf{v}}$. Esta remoção ocorre porque, quando realizamos uma intervenção em \mathbb{V} o valor desta variável é fixado e, assim, não é gerado por suas causas em \mathcal{G} .

Exemplo 4.6. Considere que $(X, Y) \in \{0, 1\}^2$ e o grafo causal é $X \rightarrow Y$. Vimos no [Exemplo 4.4](#) que o grafo causal estrutural é dado por $U_X \rightarrow X \rightarrow Y \leftarrow U_Y$. Vamos construir o grafo de resultados potenciais dadas intervenções em X . Neste caso, além dos vértices U_X, U_Y, X, Y , temos também $X_{X=0}, Y_{X=0}, X_{X=1}, Y_{X=1}$. Como não há ambiguidade neste caso, podemos abreviar os últimos quatro vértices por X_0, Y_0, X_1, Y_1 .

O grafo de resultados potenciais é ilustrado na [figur 4.1](#). O grafo causal estrutural é a reta horizontal de U_X a U_Y . Os resultados potenciais são cópias deste grafo que usam as mesmas variáveis U e em que removemos as arestas que apontam para as intervenções, X_0 e X_1 .

Uma vez definido o grafo de resultados potenciais, podemos estender a distribuição do modelo de equações estruturais para este grafo. Esta extensão envolve três etapas. Primeiramente, a distribuição de U continua a mesma. Em segundo lugar, para todo vértice do grafo de resultados potenciais, $W_{\mathbb{V}=\mathbf{v}}$, em que não houve

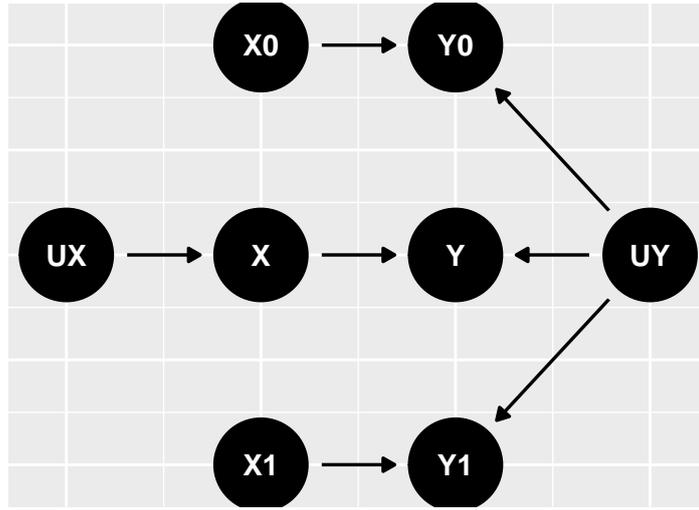


Figura 4.1.: Grafo de resultados potenciais dadas intervenções em $X \in \{0, 1\}$.

uma intervenção, este vértice é gerado pelo mesmo mecanismo que W . Isto é, $W_{\mathbb{V}=\mathbf{v}} = \mathbb{I}(g_W(U_W, Pa^*(W_{\mathbb{V}=\mathbf{v}})))$. Finalmente, se houve uma intervenção em $W_{\mathbb{V}=\mathbf{v}}$, então ela é uma variável degenerada no valor desta intervenção. Esta construção é formalizada na [Definição 4.7](#).

Definição 4.7. Seja (\mathcal{G}^+, f^+) um SCM para (\mathcal{G}, f) com funções determinísticas, g . O modelo de resultados potenciais (POM)³ dado por intervenções em \mathbf{X} , é um modelo probabilístico em um DAG, (\mathcal{G}^*, f^*) , tal que \mathcal{G}^* é o grafo de resultados potenciais dado por intervenções em \mathbf{X} ([Definição 4.5](#)) e

$$f^*(U_W) = f(U_W) \quad , \text{ para todo } W \in \mathcal{V},$$

$$f^*(W_{\mathbb{V}=\mathbf{v}} | Pa^*(W_{\mathbb{V}=\mathbf{v}})) = \begin{cases} \mathbb{I}(W_{\mathbb{V}=\mathbf{v}} = \mathbf{v}_i) & , \text{ se } W \equiv \mathbb{V}_i \\ \mathbb{I}(W_{\mathbb{V}=\mathbf{v}} = g_W(U_W, Pa^*(W_{\mathbb{V}=\mathbf{v}}))) & , \text{ caso contrário.} \end{cases}$$

O [Exemplo 4.8](#) ilustra um modelo de resultados potenciais.

Exemplo 4.8. Considere o SCM em equações estruturais em [Exemplo 4.4](#). Na construção do modelo de resultados potenciais, definimos X, Y, U_X, U_Y igualmente a em [Exemplo 4.4](#). Além disso, para cada $x > 0$, $X_x \equiv x$ e $Y_x \equiv -\log(U_Y)/X_x$.

4.1. Levando a intuição do SCM ao POM

Ainda que seja uma formalização conveniente, o POM é consideravelmente mais complexo que o SCM original. Para ganhar intuição sobre o POM, alguns lemas de tradução são fundamentais.

Lema 4.9. Se $\mathbb{V}, \mathbf{Z} \subseteq \mathcal{V}$ e $\mathbb{V} \cap \mathbf{Z} = \emptyset$, então $\mathbb{P}(\mathcal{V}_{\mathbb{V}=\mathbf{v}} = \mathcal{V}_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}} | \mathbf{Z}_{\mathbb{V}=\mathbf{v}} = \mathbf{z}) = 1$. Em particular,

$$\mathbb{P}(\mathcal{V} = \mathcal{V}_{\mathbf{Z}=\mathbf{z}} | \mathbf{Z} = \mathbf{z}) = 1.$$

O [Lema 4.9](#) conecta o dado observacional em \mathcal{V} ao resultado potencial $\mathcal{V}_{\mathbf{Z}=\mathbf{z}}$. Mais especificamente, quando observamos que $\mathbf{Z} = \mathbf{z}$, então os resultados potenciais dada a intervenção $\mathbf{Z} = \mathbf{z}$ são idênticos aos resultados

³utilizamos a sigla POM em referência ao termo em inglês “potential outcomes model”

observados. Em outras palavras, ao observarmos que $\mathbf{Z} = \mathbf{z}$, aprendemos que estamos justamente na hipótese de resultados potenciais em que $\mathbf{Z} = \mathbf{z}$.

Lema 4.10. *Se $\mathbb{V}, \mathbf{Z} \subseteq \mathcal{V}$ e $\mathbb{V} \cap \mathbf{Z} = \emptyset$, então para todo $W \in \mathcal{V}$,*

$$W_{\mathbb{V}=\mathbf{v}}\mathbb{I}(\mathbf{Z}_{\mathbb{V}=\mathbf{v}} = \mathbf{z}) \equiv W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}\mathbb{I}(\mathbf{Z}_{\mathbb{V}=\mathbf{v}} = \mathbf{z})$$

O [Lema 4.10](#) é extremamente útil, ainda que de natureza mais técnica. Ele permite que relacionemos resultados potenciais em que diferentes tipos de intervenção são adotados.

Um outro resultado essencial é o de que $\mathcal{V}_{\mathbb{V}=\mathbf{v}}$ tem a distribuição de quando realizamos a intervenção $do(\mathbb{V} = \mathbf{v})$. Esta resultado é estabelecido no [Lema 4.11](#).

Lema 4.11. *No modelo de resultados potenciais ([Definição 4.7](#)):*

$$f^*(\mathcal{V}_{\mathbb{V}=\mathbf{v}}) \equiv f(\mathcal{V}|do(\mathbb{V} = \mathbf{v})).$$

O [Lema 4.11](#) fornece uma outra forma de pensar sobre o efeito causal. Decorre do [Lema 4.11](#) que $\mathbb{E}[Y_{X=x}] = \mathbb{E}[Y|do(X = x)]$. Assim, se por exemplo X é binário, $ACE = \mathbb{E}[Y_1] - \mathbb{E}[Y_0]$. Em outras palavras, como o [Definição 4.7](#) cria variáveis aleatórias que tem a distribuição intervencional, ele permite que imaginemos o efeito causal em termos destas variáveis. Como na [capítulo 3](#) não havia acesso aos resultados potenciais, era necessário imaginar o efeito causal somente em termos da distribuição intervencional. Assim, a [Definição 4.7](#) oferece mais formas de pensar sobre o efeito causal.⁴

Uma forma alternativa de pensar sobre identificação causal está na definição de *ignorabilidade*. Dizemos que X é ignorável para medir o efeito causal em Y se ele é independente dos resultados potenciais Y_x . Em outras palavras, saber o valor de X não traz informação sobre o resultado de Y em uma outra realidade em que realizamos uma intervenção sobre X .

Definição 4.12 (Ignorabilidade). Dizemos que X é *ignorável* para medir o efeito causal em Y se $Y_x \perp^d X$.

O critério da ignorabilidade é equivalente a afirmar que X e Y não tem um ancestral comum. Em outras palavras, X é ignorável se e somente se \emptyset satisfaz o critério backdoor para medir o efeito causal de X em Y .

Lema 4.13. *As seguintes afirmações são equivalentes:*

1. \emptyset satisfaz o critério backdoor para medir o efeito causal de X em Y ,
2. $Anc(X) \cap Anc(Y) = \emptyset$, isto é, X e Y não tem um ancestral em comum, e
3. X é ignorável para medir o efeito causal em Y .

Assim, decorre do fato de que X é ignorável para o efeito causal em Y que a distribuição intervencional de Y dado X é equivalente à sua distribuição observacional. Em outras palavras, dizer que X é ignorável tem consequências similares a dizer que X é atribuído por aleatorização.

Corolário 4.14. *Se X é ignorável para medir o efeito causal em Y , então*

$$f(y|do(x)) = f(y|x).$$

⁴Esta outra forma de pensar sobre o efeito causal é tão relevante que outras construções de Inferência Causal, como o Rubin Causal Model (?) partem diretamente dela.

A *ignorabilidade condicional* oferece uma generalização da [Definição 4.12](#). Dizemos que, dado \mathbf{Z} , X é ignorável para medir o efeito causal em Y se X é independente de todo Y_x dado \mathbf{Z} .

Definição 4.15 (Ignorabilidade condicional). Dizemos que X é *condicionalmente ignorável* para medir o efeito causal em Y dado \mathbf{Z} se $Y_x \perp^d X | \mathbf{Z}$.

Se \mathbf{Z} não tem descendentes de X , a ignorabilidade condicional é uma restrição mais forte que o critério backdoor, conforme formalizado no [Lema 4.16](#).

Lema 4.16. *Suponha que $X \notin \text{Anc}(\mathbf{Z})$. Se X é condicionalmente ignorável para medir o efeito causal em Y dado \mathbf{Z} , então \mathbf{Z} satisfaz o critério backdoor para medir o efeito causal de X em Y .*

Apesar do critério backdoor e das ignorabilidade condicional não serem equivalentes, eles induzem o mesmo tipo de identificação causal.

Lema 4.17. *Se X é condicionalmente ignorável para medir o efeito causal em Y dado \mathbf{Z} , então \mathbf{Z} controla confundidores para medir o efeito causal de X em Y ([Definição 3.18](#)).*

Decorre do [Lema 4.17](#) que todas as estratégias de estimação do efeito causal estudadas na [Seção 3.2](#) também podem ser usadas sob a suposição de ignorabilidade condicional. Em outras palavras, ignorabilidade condicional fornece um critério alternativo para justificar o tipo de identificação causal obtida pelo critério backdoor.

4.1.1. Exercícios

Exercício 4.18. Prove o [Lema 4.1](#).

Exercício 4.19. Mostre que no [Exemplo 4.4](#) a distribuição de (X, Y) no SCM em equações estruturais é igual àquela no SCM original.

Exercício 4.20. Exiba um exemplo em que X é condicionalmente ignorável para Y dado \mathbf{Z} mas \mathbf{Z} não satisfaz o critério backdoor para medir o efeito causal de X em Y .

Exercício 4.21. Exiba um exemplo em que \mathbf{Z} satisfaz o critério backdoor para medir o efeito causal de X em Y mas X não é condicionalmente ignorável para Y dado \mathbf{Z} .

4.2. Variáveis Instrumentais

Há situações em que não nos sentimos confortáveis com a suposição de que observamos todos os confundidores ou todos os mediadores de X a Y . Nestes casos, não é possível justificar os métodos baseados nos critérios backdoor e frontdoor vistos na [capítulo 3](#). Variáveis instrumentais são um modo de evitar esse tipo de suposição.

Intuitivamente, uma variável instrumental, I , tem todo o seu efeito causal sobre Y mediado por X . Em outras palavras, a única forma em que I tem efeito sobre Y é na medida em que I tem efeito sobre X e, por sua vez, X tem efeito sobre Y . Por exemplo, ? estuda como participar da guerra do Vietnam, X , tem efeito sobre a renda de um indivíduo, Y . Para tal, o estudo considera os sorteios que foram realizados para determinar quem era recrutado para a guerra. O único efeito que o sorteio tem sobre a renda de um indivíduo é indireto, apenas na medida em que afeta a probabilidade de este indivíduo ir para a guerra.

Com base neste tipo de variável, sob certas circunstâncias é possível estimar o efeito causal de X em Y . A ideia básica é a de que, fazendo intervenções em I , vemos mudanças tanto em X quanto em Y . Como as mudanças em Y devem-se apenas às mudanças que ocorreram em X , pode ser possível estimar o efeito causal de X em Y .

Dada esta intuição, podemos definir formalmente uma variável instrumental. Para tal, iremos seguir de perto a abordagem em ?.

Definição 4.22. Dizemos que I é um instrumento para medir o efeito causal de X em Y se

1. I é ignorável para medir o efeito em Y .
2. $Y_{I=i, X=x} \equiv Y_{X=x}$, para todo f compatível com \mathcal{G} .
3. $Cov[I, X] \neq 0$.

Apesar de as condições na [Definição 4.22](#) terem sido utilizadas originalmente por ?, é possível reinterpretá-las utilizando o grafo causal. Já vimos no [Lema 4.13](#) que a primeira condição é equivalente a dizer que \emptyset satisfaz o critério backdoor para medir o efeito causal de I em Y . Isto é, I e Y não tem ascendentes em comum no grafo causal. Além disso, a segunda condição é equivalente a afirmar que no grafo causal todo caminho direcionado de I a Y passa por X . Isto é, X é o mediador do efeito causal de I em Y . Este resultado é apresentado no [Lema 4.23](#).

Lema 4.23. $Y_{I=i, X=x} \equiv Y_{X=x}$, para todo f compatível com \mathcal{G} se e somente se todo caminho direcionado de I a Y , C , é tal que existe j com $C_j = X$.

Sob algumas circunstâncias, a existência de um instrumento é suficiente para que seja possível identificar o efeito causal. Uma suposição usual é de que estamos analisando um CM linear Gaussiano ([Definição 2.25](#)).

Teorema 4.24. Se (\mathcal{G}, f) é um CM linear Gaussiano e I é um instrumento para medir o efeito causal de X em Y , então

$$ACE = \frac{Cov[I, Y]}{Cov[I, X]}$$

Caso o modelo causal não seja linear Gaussiano, então mais suposições são necessárias para identificar o efeito causal com base em um instrumento, Uma suposição usual é a de monotonicidade do instrumento. Segundo esta, ao aumentar o valor do instrumento por uma intervenção, o valor de X necessariamente irá aumentar

Definição 4.25. I é um instrumento monotônico para medir o efeito causal de X em Y se, para todo $i_1 > i_0$,

$$\mathbb{P}(X_{I=i_1} > X_{I=i_0}) = 1.$$

O Instrumento monotônico foi originalmente contextualizado em uma aplicação a alistados na Guerra do Vietnã ?. Pode-se imaginar que a população é dividida em 4 grupos. Pessoas que sempre iriam à guerra (always-taker), que nunca iriam à guerra (never-taker), que iriam à guerra somente se alistados (compliers), e que iriam à guerra somente se não alistados (defiers). Neste caso, o alistamento ser um instrumento monotônico corresponde a afirmar que não existem pessoas no último grupo.

Quando o instrumento é monotônico e X e I são binários, é possível identificar o efeito causal de X em Y em uma sub-população. Especificamente, é possível identificar o efeito de X em Y na sub-população em que o resultado potencial de X é diferente para cada intervenção em I . No exemplo da Guerra do Vietnã, esta é a sub-população dos *compliers*, isto é, indivíduos que iriam à guerra somente se alistados. A definição de Local Average Treatment Effect (LATE) é formalizada abaixo:

Definição 4.26. Se $X, I \in \{0, 1\}$, então

$$LATE = \mathbb{E}[Y_{X=1} - Y_{X=0} | X_{I=1} - X_{I=0} = 1].$$

O Teorema 4.27 mostra como identificar o LATE por meio de um instrumento monotônico.

Teorema 4.27. *Se $I \in \{0, 1\}$ é um instrumento monotônico para o efeito causal de $X \in \{0, 1\}$ em Y , então*

$$LATE = \frac{\mathbb{E}[Y|I = 1] - \mathbb{E}[Y|I = 0]}{\mathbb{E}[X|I = 1] - \mathbb{E}[X|I = 0]}.$$

4.3. Contrafactuais

Existem situações em que gostaríamos de saber o que teria ocorrido, caso certas condições fossem diferentes daquelas que foram efetivamente observadas. Por exemplo, considere que a perna de um indivíduo é amputada em virtude de um erro de diagnóstico médico. Neste caso, o indivíduo tem o direito a ser indenizado por seus danos. Contudo, qual o valor da indenização? Para responder a esta pergunta, somos levados a questionar como seria a vida deste indivíduo caso não houvesse o erro de diagnóstico. Este tipo de pergunta é chamada de *contrafactual*.

Uma característica fundamental de contrafactuais é que estamos interessados em uma “realidade” distinta daquela que foi observada. Contudo, se só observamos uma realidade, como é possível aprender algo sobre “realidades distintas”? Por exemplo, se supomos que não houve um erro de diagnóstico médico, o que mais seria diferente? Nesta realidade alternativa, consideramos que não houve erro médico porque há um médico muito mais concentrado, competente, ético e com exames mais precisos? Ou estamos supondo apenas que características pontuais que o levaram ao erro não estão presentes e, assim, essencialmente é o mesmo médico tratando o paciente? Ainda que não há uma resposta única para esta pergunta, há um cenário que é comumente analisado. Neste supomos que a realidade alternativa é a mais próxima possível da observada dada a restrição que um determinado fato ocorreu diferentemente.

Neste sentido, resultados potenciais são um formalismo útil. Dentro deste formalismo, consideramos que \mathcal{V} são as variáveis da “realidade observada”. Por outro lado, $\mathcal{V}_{\mathbf{X}=\mathbf{x}}$ são as variáveis que seriam observadas quando \mathbf{X} é fixado no valor \mathbf{x} . Neste formalismo (Definição 4.7), consideramos que $Y = g_Y(Pa^*(Y), U_Y)$ e $Y_{\mathbf{X}=\mathbf{x}} = g_Y(Pa^*(Y_{\mathbf{X}=\mathbf{x}}), U_Y)$. Isto é, na realidade contrafactual, $Y_{\mathbf{X}=\mathbf{x}}$ e Y são gerados pelo mesmo mecanismo, g_Y . Além disso, os ruídos locais representados por U_Y são os mesmos em Y e $Y_{\mathbf{X}=\mathbf{x}}$. Pode-se argumentar que a equivalência de mecanismos e de ruídos locais satisfaz a condição de que realidades contrafactuais devem ser tão próximas quanto possível da realidade observada.

Exemplo 4.28. Considere que X é a indicadora de que houve um erro médico e Y é a indicadora de que a perna do paciente não é amputada. Por simplicidade, vamos supor que estas são as únicas duas variáveis relevantes e que o grafo causal é \mathcal{G} tal que $X \rightarrow Y$. Também considere que $f(X = 1) = \epsilon$, $f(Y = 1|X = 1) = p_1$, e $f(Y = 1|X = 0) = p_0$, $p_0 > p_1$. Assim, o CM é (\mathcal{G}, f) .

Para definir, um modelo de resultados potenciais, é necessário determinar um SCM (Definição 4.3). Uma possibilidade é escolher $U_X, U_Y \sim U(0, 1)$, $g_X(U_X) \equiv \mathbb{I}(U_X \leq \epsilon)$, e $g_Y(U_Y, X) \equiv \mathbb{I}(U_Y \leq p_X)$. Podemos mostrar que este SCM representa o CM definido no parágrafo anterior:

$$\begin{aligned} f(X = 1) &= \mathbb{P}(\mathbb{I}(U_X \leq \epsilon)) = \epsilon && \text{Definição 4.3, } U_X \sim U(0, 1), \\ f(Y = 1|X = x) &= \mathbb{P}(\mathbb{I}(U_Y \leq p_x)) = p_x. \end{aligned}$$

Com base no modelo de resultados potenciais, podemos perguntar qual teria sido a probabilidade de que a perna do paciente não fosse amputada sem um erro médico, sabendo que observou-se o erro e a amputação:

$\mathbb{P}(Y_{X=0}|X = 1, Y = 1)$.

$$\begin{aligned}\mathbb{P}(Y_{X=0} = 1|X = 1, Y = 0) &= \mathbb{P}(\mathbb{I}(U_Y \leq p_0) = 1|\mathbb{I}(U_X \leq \epsilon) = 1, \mathbb{I}(U_Y > p_1) = 1) \\ &= \mathbb{P}(U_Y \leq p_0|U_X \leq \epsilon, U_Y > p_1) \\ &= \mathbb{P}(U_Y \leq p_0|U_Y > p_1) \\ &= \frac{\mathbb{P}(p_1 < U_Y \leq p_0)}{\mathbb{P}(U_Y > p_1)} = \frac{p_0 - p_1}{1 - p_1}\end{aligned}$$

Uma característica importante do [Exemplo 4.28](#) é que a probabilidade contrafactual depende tanto da distribuição de U_Y quanto da funções g_Y . Estas quantidades não podem ser determinadas pelos dados. Em outras palavras, as probabilidades contrafactuais dependem fundamentalmente de suposições que não podem ser testadas. No [Exemplo 4.28](#) definimos que $g_Y(U_Y, X) = \mathbb{I}(U_Y \leq p_X)$. Este acoplamento determina que todo paciente que não teve sua perna amputada com um erro médico, também não a teria sem o erro médico. Como nunca observamos ambas as situações para um mesmo paciente, esta afirmação não é testável. O [Exemplo 4.29](#) mostra que a probabilidade contrafactual varia conforme o acoplamento utilizado.

Exemplo 4.29. No [Exemplo 4.28](#), considere que $g_Y(U_Y, 1) = \mathbb{I}(U_Y \leq p_1)$ e $g_Y(U_Y, 0) = \mathbb{I}(U_Y \geq 1 - p_0)$.

$$\begin{aligned}\mathbb{P}(Y_{X=0} = 1|X = 1, Y = 0) &= \mathbb{P}(\mathbb{I}(U_Y \geq 1 - p_0) = 1|\mathbb{I}(U_X \leq \epsilon) = 1, \mathbb{I}(U_Y > p_1) = 1) \\ &= \mathbb{P}(U_Y \geq 1 - p_0|U_X \leq \epsilon, U_Y > p_1) \\ &= \mathbb{P}(U_Y > 1 - p_0|U_Y > p_1) \\ &= \frac{\mathbb{P}(U_Y > \max(1 - p_0, p_1))}{\mathbb{P}(U_Y > p_1)} = \frac{\min(p_0, 1 - p_1)}{1 - p_1}\end{aligned}$$

Neste caso, g_Y induz a mesma distribuição sobre (X, Y) que o acoplamento no [Exemplo 4.28](#). Ainda assim, a probabilidade contrafactual obtida é diferente. Isto ocorre pois, ao contrário do [Exemplo 4.28](#), a nova g_Y indica que é possível que um paciente tenha a perna amputada sem o erro médico e não a tenha com o erro.

De um ponto de vista operacional, o [Teorema 4.30](#) abaixo provê um algoritmo para calcular probabilidades contrafactuais:

Teorema 4.30 (Cálculo contrafactual).

$$\mathbb{P}(\mathbf{Y}_{\mathbf{X}=\mathbf{x}} \leq \mathbf{y}|\mathbf{Z} = \mathbf{z}) = \int \mathbb{P}(\mathbf{Y}_{\mathbf{X}=\mathbf{x}} \leq \mathbf{y}|\mathbb{U})f(\mathbb{U}|\mathbf{Z} = \mathbf{z})d\mathbb{U}$$

O [Teorema 4.30](#) indica que o cálculo de probabilidades contrafactuais pode ser dividido em duas etapas. Primeiramente, calcula-se a nova distribuição dos ruídos locais, \mathbb{U} , após aprender que $\mathbf{Z} = \mathbf{z}$, obtendo assim $f(\mathbb{U}|\mathbf{Z} = \mathbf{z})$. A seguir, calcula-se $\mathbb{P}(\mathbf{Y}_{\mathbf{X}=\mathbf{x}} \leq \mathbf{y})$ utilizando-se que a distribuição de \mathbb{U} é $f(\mathbb{U}|\mathbf{Z} = \mathbf{z})$ ao invés de $f(\mathbb{U})$.

Na próxima seção, veremos uma aplicação mais detalhada de contrafactuais no Direito. Esta discussão é baseada em ?.

4.3.1. Contrafactuais e Responsabilidade Civil

O art. 927 do Código Civil de 2002 determina que:

Aquele que, por ato ilícito, causar dano a outrem, fica obrigado a repará-lo.

Este artigo institui a Responsabilidade Aquiliana, isto é o dever daquele que causa um dano de forma ilícita a **reparar** as vítimas. Sem entrar em detalhes sobre a lei, um elemento fundamental deste conceito é a reparação.

O que é a reparação? Esta pergunta é ainda mais complicada dado que o Direito somente pode exigir que esta reparação ocorra por meio de um pagamento em dinheiro. Neste sentido, uma posição comum é definir que reparação é o valor de pagamento que torna a vítima inteira, isto é, na condição em que ela estaria caso não tivesse sofrido o dano.

Para nossos fins, a parte central desta definição de reparação é que ela exige uma consideração contrafactual. Isto é, ela questiona qual teria sido a condição da vítima caso esta não tivesse sofrido o dano. Podemos avaliar esta questão com as ferramentas desenvolvidas na [Seção 4.3](#).

Para tal, construíremos um modelo de resultados potenciais. Considere que \mathcal{V} é o conjunto de todas as variáveis juridicamente relevantes. Dentre elas, I é a indicadora de que o causador do dano agiu de forma ilícita. Além disso, $E \subseteq \mathcal{V}$ é um conjunto de variáveis que foi observada e trazida como evidência à Justiça. Finalmente, $U(\mathcal{V}, m)$ é uma função de utilidade que indica o quão desejável é para a vítima obter o resultado \mathcal{V} e, além disso, uma indenização monetária de m . Dentro deste contexto, uma quantificação do dano é uma função, $Q(E) \in \mathbb{R}^+$, que define uma compensação para cada possível evidência apresentada.

Dentro deste contexto, uma primeira pergunta é o quanto da evidência pode ser usada para quantificar o dano. Neste aspecto, há uma tensão importante entre previsibilidade e reparação (?). Considere que um diário com a assinatura de Janis Joplin é destruído em um sebo antes da fama da cantora. Poderia a posterior fama de Joplin ser levada em consideração na reparação? De forma mais geral, quanto da evidência conhecida no momento do julgamento pode ser usada na quantificação do dano?

De uma perspectiva social, cada possível resposta apresenta vantagens e desvantagens. Por um lado, o uso de muita informação permite que seja possível colocar a vítima no presente o mais próximo da condição em que ela estaria caso o ato ilícito não tivesse ocorrido. Por outro lado, quanto mais informação é usada, mais imprevisível se torna o valor da reparação no momento em que o ilícito é cometido.

Formalmente, podemos imaginar ao menos três respostas para a pergunta levantada. A primeira resposta é de que toda evidência pode ser usada. Isto é, não há restrições sobre Q . A segunda resposta é de que nenhuma evidência pode ser usada, isto é, Q deve ser uma função constante. Finalmente, uma resposta intermediária é obtida supondo que Q somente leva em consideração se a situação da vítima foi superior ou inferior àquela que ela poderia esperar sem o ilícito. Formalmente, tomando $g(E) = \mathbb{I}(\mathbb{E}[U(\mathcal{V}, 0)|E] > \mathbb{E}[U(\mathcal{V}_{X=0}, 0)])$, temos que se $g(E_1) = g(E_2)$, então $Q(E_1) = Q(E_2)$. Em última análise, a escolha da resposta correta não é uma questão científica, mas de Direito. Contudo, utilizando contrafactuais podemos analisá-la e indicar os aspectos fundamentais envolvidos na escolha.

Um outro aspecto fundamental na quantificação de danos é a especificação de Q . Para tal, há duas possíveis interpretações de reparação. Por um lado, pode-se entender que a reparação deve levar a vítima o mais próximo possível ao estado em que ela estaria sem o ato ilícito, levando em consideração toda a evidência disponível. Uma possível formalização deste raciocínio é:

$$Q = \arg \min_{Q^*} \mathbb{E}[(U(\mathcal{V}, Q^*(E)) - U(\mathcal{V}_{X=0}, 0))^2]$$

Por outro lado, pode-se entender que, no momento em que o ato ilícito foi cometido, a vítima deveria estar indiferente entre não sofrer o ilícito ou sofrê-lo e receber a indenização. Esta outra interpretação pode ser formalizada

da seguinte forma:

$$Q = \arg \min_{Q^*: \mathbb{E}[U(\mathcal{V}, Q^*)] = \mathbb{E}[U(\mathcal{V}_{X=0}, 0)]} \mathbb{E}[(U(\mathcal{V}, Q^*(E)) - U(\mathcal{V}_{X=0}, 0))^2]$$

Finalmente, como estabelecer um modelo de resultados potenciais a partir de um modelo causal? Como vimos anteriormente, um mesmo CM pode ser compatível com diversos SCM. Além disso, cada SCM pode levar a avaliações distintas de contrafactuais. Uma possível resposta é que o SCM deve ser estabelecido pelo perito. Contudo, o custo de um perito pode ser incompatível com o valor do pedido. Neste caso, algumas alternativas são supor que \mathcal{V} e $\mathcal{V}_{X=0}$ são independentes ou tomar a sua dependência de forma a minimizar $\mathbb{E}[(U(\mathcal{V}, 0) - U(\mathcal{V}_{X=0}, 0))^2]$.

4.3.2. Exercícios

Exercício 4.31. Com base na discussão levantada na [Seção 4.3.1](#), discuta formas de reparação para o caso levantado no [Exemplo 4.28](#).

Exercício 4.32. Suponha que sem o ato ilícito, um paciente tem o tempo de sobrevivência dado pela acumulada F_0 . Por outro lado, com o ato ilícito, seu tempo de sobrevivência é dado por F_1 . Ocorrido o ilícito, o paciente falece após 1 ano. Discuta possíveis reparações para o paciente com base na [Seção 4.3.1](#).

5. Descoberta Causal

Há situações em que observamos dados de um CM desconhecido, (\mathcal{G}, f) . Nestes casos, podemos ter interesse em estimar o grafo causal, \mathcal{G} . Neste capítulo, estudaremos estratégias para esta estimação.

Antes de apresentar métodos de estimação, avaliaremos condições para que o problema de estimação esteja bem posto. Especificamente, a próxima seção mostra que, de forma irrestrita, o grafo causal é estatisticamente não-identificável. Assim, comumente precisamos nos contentar com um objetivo menos ambicioso de estimação.

5.1. Identificabilidade na Descoberta Causal

Para contextualizar o problema da identificabilidade na Descoberta Causal, podemos começar com um exemplo simplista. Considere que observamos as variáveis: $\mathcal{V} = \{X, Y\}$. Neste caso, há pelo menos duas dificuldades.

Primeiramente, considere que f é tal que $X \perp\!\!\!\perp^f Y$. Neste caso, apesar de intuitivamente considerarmos o grafo causal, $X \rightarrow Y$, a densidade f também é compatível com $X \rightarrow Y$. De fato, $X \rightarrow Y$ determina que $f(x, y) = f(x)f(y|x)$, o que é verdadeiro mesmo quando X e Y são independentes.

Contudo, ainda que ambos os grafos sejam compatíveis com f , o grafo $X \rightarrow Y$ é mais informativo. Somente com base neste grafo podemos deduzir que X é independente de Y . Essa constatação elucida que não buscamos apenas um grafo causal, \mathcal{G} , compatível com a distribuição geradora dos dados, f . Desejamos que \mathcal{G} permita deduzir o maior número possível de relações de independência condicional presentes em f . Uma maneira de lidar com este problema é trocar o objetivo de encontrar \mathcal{G} compatível com f por encontrar \mathcal{G} **fiel** a f :

Definição 5.1. Dizemos que f é fiel a \mathcal{G} ou, equivalentemente, que \mathcal{G} é fiel a f , se para quaisquer $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$, $\mathbf{X} \perp\!\!\!\perp^f \mathbf{Y}|\mathbf{Z}$ se e somente se $\mathbf{X} \perp\!\!\!\perp^{\mathcal{G}} \mathbf{Y}|\mathbf{Z}$.

Podemos interpretar a [Definição 5.1](#) à luz do [Teorema 2.49](#). Se \mathcal{G} é compatível com f e $\mathbf{X} \perp\!\!\!\perp^{\mathcal{G}} \mathbf{Y}|\mathbf{Z}$, então o [Teorema 2.49](#) indica que $\mathbf{X} \perp\!\!\!\perp^f \mathbf{Y}|\mathbf{Z}$. Contudo, em geral não é possível inferir uma d-separação em um grafo compatível a partir de uma relação de independência condicional. Quando \mathcal{G} é fiel a f , toda relação de independência condicional implica uma d-separação em \mathcal{G} . Isto é, \mathcal{G} traz a maior informação possível sobre as relações de independência condicional em f .

O segundo problema de identificabilidade é que pode haver mais de um grafo causal fiel à distribuição geradora dos dados. Como ilustração, considere novamente que $\mathcal{V} = \{X, Y\}$ e que f é tal que X e Y não são independentes. Neste caso, tanto $X \rightarrow Y$ quanto $X \leftarrow Y$ são fiéis a f . Em outras palavras, se não fizermos mais suposições, ambos os grafos explicam igualmente bem f .

Com base nesta observação, podemos definir um tipo de não-identificabilidade em descoberta causal. Se dois grafos causais são fiéis às mesmas distribuições, então não importa qual seja a densidade geradora dos dados, não é possível diferenciá-los por meio dos dados. Neste caso, dizemos que os grafos são **fielmente indistinguíveis**:

Definição 5.2. Dois grafos, \mathcal{G} e \mathcal{G}^* , são fielmente indistinguíveis ($\mathcal{G} \sim \mathcal{G}^*$) se, para todo f ,

$$f \text{ é fiel a } \mathcal{G} \text{ se e somente se } f \text{ é fiel a } \mathcal{G}^*.$$

A seguir, a [Definição 5.4](#) e o [Teorema 5.5](#) mostram uma operacionalização da relação de indistinguibilidade fiel.

Definição 5.3. Dizemos que (V_1, V_2, V_3) formam um colisor desprotegido em \mathcal{G} se $V_1 \xrightarrow{\mathcal{G}} V_2 \xleftarrow{\mathcal{G}} V_3$ e V_1 e V_3 não são adjacentes em \mathcal{G} .

Definição 5.4. Os DAGs \mathcal{G} e \mathcal{G}^* tem o mesmo padrão se:

1. Para quaisquer vértices V_1 e V_2 , V_1 e V_2 são adjacentes em \mathcal{G} se e somente se V_1 e V_2 são adjacentes em \mathcal{G}^* .
2. (V_1, V_2, V_3) é um colisor desprotegido em \mathcal{G} se e somente se é um colisor desprotegido em \mathcal{G}^* .

Teorema 5.5 (?). $\mathcal{G} \sim \mathcal{G}^*$ se e somente se \mathcal{G} e \mathcal{G}^* tem o mesmo padrão.

De forma geral, o [Teorema 5.5](#) mostra que, sem maiores suposições, só é possível aprender o padrão de um grafo a partir dos dados. Isto é, é possível determinar todas as relações de adjacência e, também, os colisores desprotegidos existentes no grafo. A seguir, estudaremos algoritmos que se propõem a descobrir esse padrão.

5.2. Algoritmos de Descoberta Causal

5.2.1. Algoritmos baseados em testes de independência condicional

Algoritmo SGS

Algoritmo 5.6. Algoritmo SGS

1. Inicie um grafo não direcionado completo em \mathcal{V} .
2. Para cada $V_1, V_2 \in \mathcal{V}$, se $H_0 : V_1 \perp\!\!\!\perp V_2 | S$ é rejeitada para algum $S \subseteq \mathcal{V} - \{V_1, V_2\}$, **retire** a aresta $\{V_1, V_2\}$.
3. Para todos $V_1, V_2, V_3 \in \mathcal{V}$ tais que $\{V_1, V_2\}$ e $\{V_2, V_3\}$ são adjacentes, mas $\{V_1, V_3\}$ não o são, caso exista $S \subseteq \{V_2\} \cup \mathcal{V} - \{V_1, V_3\}$ tal que $H_0 : V_1 \perp\!\!\!\perp V_2 | S$ não é rejeitada, **direcione** as arestas $V_1 \rightarrow V_2 \leftarrow V_3$.
4. Repetir até que não haja mudanças:
 - (a) Se $A \rightarrow B$, B e C são adjacentes e não direcionados e A e C não são adjacentes, **direcione** $B \rightarrow C$.
 - (b) Se B é descendente de A e $\{A, B\}$ são adjacentes, **direcione** $A \rightarrow B$.

Algoritmo PC

Definição 5.7. $Adj_{\mathcal{G}}(X)$ é o conjunto de vértices em \mathcal{V} que é adjacente a X em \mathcal{G} .

Algoritmo 5.8. Algoritmo PC

1. Inicie um grafo não direcionado completo em \mathcal{V} , \mathcal{G} .
2. Inicialize $n = 0$.
3. Enquanto houver um vértice $X \in \mathcal{V}$ com $|Adj_{\mathcal{G}}(X)| > n + 1$:
 - (a) Para cada $X \in \mathcal{V}$ com $|Adj_{\mathcal{G}}(X)| \geq n + 1$, cada $Y \in Adj_{\mathcal{G}}(X)$ e $S \subseteq Adj_{\mathcal{G}}(X) - \{Y\}$ em que $|S| = n$, se $H_0 : X \perp\!\!\!\perp Y | S$ é rejeitada, então **retire** a aresta $\{X, Y\}$ de \mathcal{G} e defina $Sep(\{X, Y\}) = S$.
 - (b) Defina $n = n + 1$.

4. Para cada $X, Y, Z \in \mathcal{V}$ tais que $\{X, Y\}$ e $\{Y, Z\}$ são adjacentes, mas $\{X, Z\}$ não o são e $Y \notin \text{Sep}(\{X, Z\})$, **oriente** $X \rightarrow Y \leftarrow Z$.
5. Repetir até que não haja mudanças:
 - (a) Se $A \rightarrow B$, B e C são adjacentes e não direcionados e A e C não são adjacentes, **direcione** $B \rightarrow C$.
 - (b) Se B é descendente de A e $\{A, B\}$ são adjacentes, **direcione** $A \rightarrow B$.

A. Demonstrações

A.1. Seção 2.1.4 (Exemplos de Modelo Probabilístico em um DAG)

Prova do Lema 2.18.

$$\begin{aligned} f(v_1, v_3|v_2) &= \frac{f(v_1, v_2, v_3)}{f(v_2)} \\ &= \frac{f(v_2)f(v_1|v_2)f(v_3|v_2)}{f(v_2)} && \text{Definição 2.11} \\ &= f(v_1|v_2)f(v_3|v_2) \end{aligned}$$

□

Prova do Lema 2.19. Considere que $V_2 \sim \text{Bernoulli}(0.02)$. Além disso, $V_1, V_3 \in \{0, 1\}$ são independentes dado V_2 . Também, $\mathbb{P}(V_1 = 1|V_2 = 1) = \mathbb{P}(V_3 = 1|V_2 = 1) = 0.9$ e $\mathbb{P}(V_1 = 1|V_2 = 0) = \mathbb{P}(V_3 = 1|V_2 = 0) = 0.05$. Note que, por construção, \mathbb{P} é compatível com [figur 2.2](#). Isto é, $P(v_1, v_2, v_3) = \mathbb{P}(v_2)\mathbb{P}(v_1|v_2)\mathbb{P}(v_3|v_2)$. Além disso,

$$\begin{aligned} \mathbb{P}(V_1 = 1) &= \mathbb{P}(V_1 = 1, V_2 = 1) + \mathbb{P}(V_1 = 1, V_2 = 0) \\ &= \mathbb{P}(V_2 = 1)\mathbb{P}(V_1 = 1|V_2 = 1) + \mathbb{P}(V_2 = 0)\mathbb{P}(V_1 = 1|V_2 = 0) \\ &= 0.02 \cdot 0.9 + 0.98 \cdot 0.05 = 0.067 \end{aligned}$$

Por simetria, $\mathbb{P}(V_3 = 1) = 0.067$. Além disso,

$$\begin{aligned} \mathbb{P}(V_1 = 1, V_3 = 1) &= \mathbb{P}(V_1 = 1, V_3 = 1, V_2 = 1) + \mathbb{P}(V_1 = 1, V_3 = 1, V_2 = 0) \\ &= \mathbb{P}(V_2 = 1)\mathbb{P}(V_1 = 1|V_2 = 1)\mathbb{P}(V_3 = 1|V_2 = 1) + \mathbb{P}(V_2 = 0)\mathbb{P}(V_1 = 1|V_2 = 0)\mathbb{P}(V_3 = 1|V_2 = 0) \\ &= 0.02 \cdot 0.9 \cdot 0.9 + 0.98 \cdot 0.05 \cdot 0.05 = 0.01865 \end{aligned}$$

Como $\mathbb{P}(V_1 = 1)\mathbb{P}(V_3 = 1) = 0.067 \cdot 0.067 \approx 0.0045 \neq 0.01865 = \mathbb{P}(V_1 = 1, V_3 = 1)$, temos que V_1 e V_3 não são independentes. □

Prova do Lema 2.20.

$$\begin{aligned} f(v_3|v_1, v_2) &= \frac{f(v_1, v_2, v_3)}{f(v_1, v_2)} \\ &= \frac{f(v_1)f(v_2|v_1)f(v_3|v_2)}{f(v_1)f(v_2|v_1)} && \text{Definição 2.11} \\ &= f(v_3|v_2) \end{aligned}$$

□

Prova do Lema 2.21. Considere que $V_1 \sim \text{Bernoulli}(0.5)$, $\mathbb{P}(V_2 = 1|V_1 = 1) = 0.9$, $\mathbb{P}(V_2 = 1|V_1 = 0) = 0.05$, $\mathbb{P}(V_3 = 1|V_2 = 1, V_1) = 0.9$, e $\mathbb{P}(V_3 = 1|V_2 = 0, V_1) = 0.05$. Note que (V_1, V_2, V_3) formam uma Cadeia de Markov. Note que, por construção, \mathbb{P} é compatível com [figur 2.3](#). Isto é, $P(v_1, v_2, v_3) = \mathbb{P}(v_1)\mathbb{P}(v_2|v_1)\mathbb{P}(v_3|v_2)$. Além disso,

$$\begin{aligned}\mathbb{P}(V_3 = 1) &= \mathbb{P}(V_1 = 0, V_2 = 0, V_3 = 1) + \mathbb{P}(V_1 = 0, V_2 = 1, V_3 = 1) \\ &\quad + \mathbb{P}(V_1 = 1, V_2 = 0, V_3 = 1) + \mathbb{P}(V_1 = 1, V_2 = 1, V_3 = 1) \\ &= 0.5 \cdot 0.9 \cdot 0.05 + 0.5 \cdot 0.05 \cdot 0.9 \\ &\quad + 0.5 \cdot 0.05 \cdot 0.05 + 0.5 \cdot 0.9 \cdot 0.9 = 0.45125\end{aligned}$$

Além disso,

$$\begin{aligned}\mathbb{P}(V_1 = 1, V_3 = 1) &= \mathbb{P}(V_1 = 1, V_2 = 0, V_3 = 1) + \mathbb{P}(V_1 = 1, V_2 = 1, V_3 = 1) \\ &= 0.5 \cdot 0.05 \cdot 0.9 + 0.5 \cdot 0.9 \cdot 0.9 = 0.40625\end{aligned}$$

Como $\mathbb{P}(V_1 = 1)\mathbb{P}(V_3 = 1) = 0.5 \cdot 0.45125 \approx 0.226 \neq 0.40625 = \mathbb{P}(V_1 = 1, V_3 = 1)$, temos que V_1 e V_3 não são independentes. \square

Prova do Lema 2.22.

$$\begin{aligned}f(v_1, v_3) &= \int f(v_1, v_2, v_3)dv_2 \\ &= \int f(v_1)f(v_3)f(v_2|v_1, v_3)dv_2 && \text{Definição 2.11} \\ &= f(v_1)f(v_3) \int f(v_2|v_1, v_3)dv_2 \\ &= f(v_1)f(v_3)\end{aligned}$$

\square

Prova do Lema 2.23. Considere que V_1 e V_3 são independentes e tem distribuição Bernoulli(0.5). Além disso, $V_2 \equiv V_1 + V_3$. Como $\mathbb{P}(V_3 = 1) = 0.5$ e $\mathbb{P}(V_3 = 1|V_1 = 1, V_2 = 2) = 1$, conclua que $V_1 \not\perp V_3|V_2$. \square

A.2. Seção 2.1.5 (Modelo Causal (Causal Model))

Prova do Lema 2.27. Realizaremos a demonstração por indução. Para tal, defina $\mathcal{V}^{(1)} = \{V \in \mathcal{V} : Pa(V) = \emptyset\}$, e para cada $i > 2$, $\mathcal{V}^{(i)} = \{V \in \mathcal{V} : Pa(V) \subseteq \mathcal{V}^{(i-1)}\}$.

Se $Y \in \mathcal{V}^{(1)}$, então por construção $\mathbb{E}[Y] = \mu_Y$. Também, como $Pa(Y) = \emptyset$, $\mathbb{C}_{V,Y} = \emptyset$, para todo $V \neq Y$, e $\mathbb{C}_{Y,Y}$ tem apenas o caminho unitário, $C^* = (Y)$. Assim,

$$\begin{aligned}\sum_{V \in \mathcal{V}} \sum_{C \in \mathbb{C}_{V,Y}} \mu_V \cdot \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i} &= \mu_Y \cdot \prod_{i=1}^{|C^*|-1} \beta_{C_{i+1}^*, C_i^*} \\ &= \mu_Y\end{aligned}$$

A seguir, suponha que para todo $W \in \mathcal{V}^{(i-1)}$, $\mathbb{E}[W] = \sum_{V \in \mathcal{V}} \sum_{C \in \mathbb{C}_{V,W}} \mu_V \cdot \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i}$ e tome $Y \in \mathcal{V}^{(i)}$. Como todo caminho direcionado que chega em Y a partir de $V \neq Y$ tem como penúltimo elemento um pai de Y ,

podemos escrever

$$\mathbb{C}_{V,Y} = \cup_{W \in Pa(Y)} \{(C, Y) : C \in \mathbb{C}_{V,W}\} \quad (\text{A.1})$$

Portanto, obtemos:

$$\begin{aligned} & \sum_{V \in \mathcal{V}} \sum_{C \in \mathbb{C}_{V,Y}} \mu_V \cdot \prod_{i=1}^{|\mathbb{C}|-1} \beta_{C_{i+1}, C_i} \\ &= \mu_Y + \sum_{V \neq Y} \sum_{\cup_{W \in Pa(Y)} \{C^* = (C, Y) : C \in \mathbb{C}_{V,W}\}} \mu_V \cdot \left(\prod_{i=1}^{|\mathbb{C}^*|-1} \beta_{C_{i+1}^*, C_i^*} \right) && \text{likning (A.1)} \\ &= \mu_Y + \sum_{V \neq Y} \sum_{W \in Pa(Y)} \sum_{C \in \mathbb{C}_{V,W}} \mu_V \cdot \left(\prod_{i=1}^{|\mathbb{C}|-1} \beta_{C_{i+1}, C_i} \right) \cdot \beta_{Y,W} \\ &= \mu_Y + \sum_{W \in Pa(Y)} \beta_{Y,W} \sum_{V \in \mathcal{V}} \sum_{C \in \mathbb{C}_{V,W}} \mu_V \cdot \left(\prod_{i=1}^{|\mathbb{C}|-1} \beta_{C_{i+1}, C_i} \right) \\ &= \mu_Y + \sum_{W \in Pa(Y)} \beta_{Y,W} \mathbb{E}[W] && W \in \mathcal{V}^{(i-1)} \\ &= \mathbb{E} \left[\mu_Y + \sum_{W \in Pa(Y)} \beta_{Y,W} W \right] \\ &= \mathbb{E}[\mathbb{E}[Y | Pa(Y)]] = \mathbb{E}[Y] && \text{Definição 2.25} \end{aligned}$$

□

A.3. Seção 2.2 (Independência Condicional e D-separação)

A.3.1. Relativas ao Lema 2.45

Prova do Lema 2.45. A prova consistirá em demonstrar que, para cada i , a afirmação i decorre da afirmação $i - 1$. Finalmente, a afirmação 1 decorre da afirmação 4. Os símbolos \mathbf{X} e \mathbf{x} referem-se a $(\mathbf{X}_1, \dots, \mathbf{X}_d)$ e $(\mathbf{x}_1, \dots, \mathbf{x}_d)$.

- (1 \implies 2)

$$\begin{aligned} f(\mathbf{x}|\mathbf{y}) &= \prod_{j=1}^d f(\mathbf{x}_j|\mathbf{y}) && (1) \\ &= \prod_{j=1}^d h(\mathbf{x}_j, \mathbf{y}) && h(\mathbf{x}_j, \mathbf{y}) = f(\mathbf{x}_j|\mathbf{y}) \end{aligned}$$

- (2 \implies 3) Note que,

$$\begin{aligned}
f(\mathbf{x}_i | \mathbf{x}_{-i}, \mathbf{y}) &= \frac{f(\mathbf{x} | \mathbf{y})}{f(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_d | \mathbf{y})} \\
&= \frac{f(\mathbf{x} | \mathbf{y})}{\int_{\mathbb{R}} f(\mathbf{x} | \mathbf{y}) d\mathbf{x}_i} \\
&= \frac{\prod_{j=1}^d h_j(\mathbf{x}_j, \mathbf{y})}{\int_{\mathbb{R}} \prod_{j=1}^d h_j(\mathbf{x}_j, \mathbf{y}) d\mathbf{x}_i} \quad (2) \\
&= \frac{\prod_{j=1}^d h_j(\mathbf{x}_j, \mathbf{y})}{\prod_{j \neq i} h_j(\mathbf{x}_j, \mathbf{y}) \int_{\mathbb{R}} h_i(\mathbf{x}_i, \mathbf{y}) d\mathbf{x}_i} \\
&= \frac{\tilde{h}_i(\mathbf{x}_i, \mathbf{y})}{\int_{\mathbb{R}} h_i(\mathbf{x}_i, \mathbf{y}) d\mathbf{x}_i} \\
&= \frac{\prod_{j \neq i} \int_{\mathbb{R}} h_j(\mathbf{x}_j, \mathbf{y}) d\mathbf{x}_j}{\prod_{j \neq i} \int_{\mathbb{R}} h_j(\mathbf{x}_j, \mathbf{y}) d\mathbf{x}_j} \cdot \frac{h_i(\mathbf{x}_i, \mathbf{y})}{\int_{\mathbb{R}} h_i(\mathbf{x}_i, \mathbf{y}) d\mathbf{x}_i} \\
&= \frac{\int_{\mathbb{R}^{d-1}} \prod_{j=1}^d h_j(\mathbf{x}_j, \mathbf{y}) d\mathbf{x}_{-i}}{\int_{\mathbb{R}^d} \prod_{j=1}^d h_j(\mathbf{x}_j, \mathbf{y}) d\mathbf{x}} \\
&= \frac{\int_{\mathbb{R}^{d-1}} f(\mathbf{x} | \mathbf{y}) d\mathbf{x}_{-i}}{\int_{\mathbb{R}^d} f(\mathbf{x} | \mathbf{y}) d\mathbf{x}} \quad (2) \\
&= f(\mathbf{x}_i | \mathbf{y})
\end{aligned}$$

- (3 \implies 4)

$$\begin{aligned}
f(\mathbf{x}_i | \mathbf{x}_1^{i-1}, \mathbf{y}) &= \frac{f(\mathbf{x}_1^i | \mathbf{y})}{f(\mathbf{x}_1^{i-1} | \mathbf{y})} \\
&= \frac{\int_{\mathbb{R}^{d-i}} f(\mathbf{x} | \mathbf{y}) d\mathbf{x}_{i+1}^d}{f(\mathbf{x}_1^{i-1} | \mathbf{y})} \\
&= \frac{\int_{\mathbb{R}^{d-i}} f(\mathbf{x}_{-i} | \mathbf{y}) f(\mathbf{x}_i | \mathbf{x}_{-i}, \mathbf{y}) d\mathbf{x}_{i+1}^d}{f(\mathbf{x}_1^{i-1} | \mathbf{y})} \\
&= \frac{f(\mathbf{x}_i | \mathbf{y}) \int_{\mathbb{R}^{d-i}} f(\mathbf{x}_{-i} | \mathbf{y}) d\mathbf{x}_{i+1}^d}{f(\mathbf{x}_1^{i-1} | \mathbf{y})} \quad (3) \\
&= \frac{f(\mathbf{x}_i | \mathbf{y}) f(\mathbf{x}_1^{i-1} | \mathbf{y})}{f(\mathbf{x}_1^{i-1} | \mathbf{y})} \\
&= f(\mathbf{x}_i | \mathbf{y})
\end{aligned}$$

- (4 \implies 1)

$$\begin{aligned}
f(\mathbf{x} | \mathbf{y}) &= \prod_{i=1}^d f(\mathbf{x}_i | \mathbf{x}_1^{i-1}, \mathbf{y}) \\
&= \prod_{i=1}^d f(\mathbf{x}_i | \mathbf{y}) \quad (4)
\end{aligned}$$

□

A.3.2. Teorema 2.49

Lema A.1. *Seja $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ um DAG. Se $\mathcal{A} = \mathbb{V}_1 \cup \mathbb{V}_2 \cup \mathbb{V}_3$ é ancestral e $\mathbb{V}_1 \perp \mathbb{V}_2 | \mathbb{V}_3$, então, para todo f compatível com \mathcal{G} , $\mathbb{V}_1 \perp^f \mathbb{V}_2 | \mathbb{V}_3$.*

Demonstração. Defina $\mathbb{V}_1^* = \{V \in \mathcal{A} : V \in \mathbb{V}_1 \text{ ou } V_1 \rightarrow V, \text{ para algum } V_1 \in \mathbb{V}_1\}$ e $\mathbb{V}_2^* = \mathcal{A} - \mathbb{V}_1^*$. Como $\mathbb{V}_1 \perp \mathbb{V}_2 | \mathbb{V}_3$, decorre de [Definição 2.48](#) que não existe $V_1 \in \mathbb{V}_1$ e $V_2 \in \mathbb{V}_2$ tal que $V_1 \rightarrow V_2$. Portanto,

$$\mathbb{V}_1^* \subseteq \mathbb{V}_1 \cup \mathbb{V}_3 \text{ e } \mathbb{V}_2^* \subseteq \mathbb{V}_2 \cup \mathbb{V}_3 \quad (\text{A.2})$$

A seguir, demonstraremos que

$$\forall i \in \{1, 2\} \text{ e } V_i^* \in \mathbb{V}_i^* : Pa(V_i^*) \subseteq \mathbb{V}_i \cup \mathbb{V}_3 \quad (\text{A.3})$$

Tome $V_1^* \in \mathbb{V}_1^*$. Como $V_1^* \in \mathcal{A}$ e \mathcal{A} é ancestral, decorre da [Definição 2.9](#) que $Pa(V_1^*) \subseteq \mathcal{A}$. Assim, basta demonstrar que $Pa(V_1^*) \cap \mathbb{V}_2 = \emptyset$. Se $V_1^* \in \mathbb{V}_1$, então decorre de [Definição 2.48](#) que não existe $V_2 \in \mathbb{V}_2$ tal que $V_2 \rightarrow V_1^*$. Caso contrário, se $V_1^* \in \mathbb{V}_3$, então existe $V_1 \in \mathbb{V}_1$ tal que $V_1 \rightarrow V_1^*$. Decorre de [Definição 2.48](#) que não existe $V_1 \in \mathbb{V}_1$, $V_2 \in \mathbb{V}_2$ e $V_3 \in \mathbb{V}_3$ tais que V_3 é um colisor entre V_1 e V_2 , isto é, $V_1 \rightarrow V_3 \leftarrow V_2$. Portanto, não existe $V_2 \in \mathbb{V}_2$ tal que $V_2 \rightarrow V_1^*$. Conclua que $Pa(V_1^*) \subseteq \mathbb{V}_1 \cup \mathbb{V}_3$.

A seguir, note que pela definição de \mathbb{V}_1^* , se $V \in \mathcal{A}$ é tal que existe $V_1 \in \mathbb{V}_1$ com $V_1 \rightarrow V$, então $V \in \mathbb{V}_1^*$. Portanto, como $\mathbb{V}_2^* = \mathcal{V} - \mathbb{V}_1^*$, para todo $V_2^* \in \mathbb{V}_2^*$, não existe $V_1 \in \mathbb{V}_1$ tal que $V_1 \rightarrow V_2^*$. Isto é, $Pa(V_2^*) \subseteq \mathcal{V} - \mathbb{V}_1$. Como $V_2^* \in \mathcal{A}$ e \mathcal{A} é ancestral, conclua da [Definição 2.9](#) que $Pa(V_2^*) \subseteq \mathcal{A}$. Combinando as duas últimas frases, $Pa(V_2^*) \subseteq \mathbb{V}_2 \cup \mathbb{V}_3$.

Decorre da conclusão dos dois últimos parágrafos que [likning \(A.3\)](#) está demonstrado.

$$\begin{aligned} f(\mathbb{V}_1, \mathbb{V}_2 | \mathbb{V}_3) &= \frac{f(\mathbb{V}_1, \mathbb{V}_2, \mathbb{V}_3)}{f(\mathbb{V}_3)} \\ &= \frac{\prod_{V \in \mathcal{A}} f(V | Pa(V))}{f(\mathbb{V}_3)} && \text{Lema 2.17} \\ &= \frac{\left(\prod_{V_1^* \in \mathbb{V}_1^*} f(V_1^* | Pa(V_1^*)) \right) \left(\prod_{V_2^* \in \mathbb{V}_2^*} f(V_2^* | Pa(V_2^*)) \right)}{f(\mathbb{V}_3)} && \mathbb{V}_1^* \text{ e } \mathbb{V}_2^* \text{ particionam } \mathcal{A} \\ &= h_1(\mathbb{V}_1, \mathbb{V}_3) h_2(\mathbb{V}_2, \mathbb{V}_3) && \text{likningene (A.2) e (A.3)} \end{aligned}$$

Assim, decorre do [Lema 2.45](#) que $\mathbb{V}_1 \perp^f \mathbb{V}_2 | \mathbb{V}_3$. □

Lema A.2. *Se f é compatível com \mathcal{G} e $\mathbb{V}_1 \perp \mathbb{V}_2 | \mathbb{V}_3$, então $\mathbb{V}_1 \perp^f \mathbb{V}_2 | \mathbb{V}_3$.*

Demonstração. Defina $\mathcal{A} = Anc(\mathbb{V}_1 \cup \mathbb{V}_2 \cup \mathbb{V}_3)$, $\mathbb{V}_1^* = \{V \in \mathcal{A} : V \text{ não é d-separado de } \mathbb{V}_1 | \mathbb{V}_3\}$, e $\mathbb{V}_2^* = \mathcal{A} - \mathbb{V}_1^*$. Por definição,

$$\mathbb{V}_1 \subseteq \mathbb{V}_1^* \text{ e } \mathbb{V}_2 \subseteq \mathbb{V}_2^* \quad (\text{A.4})$$

O primeiro é provar que $\mathbb{V}_1^* \perp \mathbb{V}_2^* | \mathbb{V}_3$. Pela definição de \mathbb{V}_2^* , para todo $V_1 \in \mathbb{V}_1$ e $V_2^* \in \mathbb{V}_2^*$, $V_1 \perp V_2^* | \mathbb{V}_3$, isto é,

$$\mathbb{V}_1 \perp \mathbb{V}_2^* | \mathbb{V}_3 \quad (\text{A.5})$$

Suponha por absurdo que existam $V_1^* \in \mathbb{V}_1^*$ e $V_2^* \in \mathbb{V}_2^*$ tais que V_1^* e V_2^* não são d-separados dado \mathbb{V}_3 . Portanto, existe um caminho ativo dado \mathbb{V}_3 , $(V_1^*, C_2, \dots, C_{n-1}, V_2^*)$. Pela definição de \mathbb{V}_1^* , existe $V_1 \in \mathbb{V}_1$ e um caminho ativo

dado $\mathbb{V}_3, (V_1, C_2^*, \dots, C_{m-1}^*, V_1^*)$. Assim, $(V_1, C_2^*, \dots, C_{m-1}^*, V_1^*, C_2, \dots, C_{n-1}, V_2^*)$ é um caminho ativo dado \mathbb{V}_3 de V_1 a V_2^* , uma contradição com [likning \(A.5\)](#). Conclua que $\mathbb{V}_1 \perp \mathbb{V}_2^* | \mathbb{V}_3$.

A seguir, provaremos que $\mathbb{V}_1^* \perp^f \mathbb{V}_2^* | \mathbb{V}_3$. Como $\mathcal{A} = \text{Anc}(\mathbb{V}_1 \cup \mathbb{V}_2 \cup \mathbb{V}_3)$, decorre do [Lema 2.10](#) que \mathcal{A} é ancestral. Portanto, como $\mathcal{A} = \mathbb{V}_1^* \cup \mathbb{V}_2^* \cup \mathbb{V}_3$ e $\mathbb{V}_1^* \perp \mathbb{V}_2^* | \mathbb{V}_3$, decorre do [Lema A.1](#) que $\mathbb{V}_1^* \perp^f \mathbb{V}_2^* | \mathbb{V}_3$.

Como $\mathbb{V}_1^* \perp^f \mathbb{V}_2^* | \mathbb{V}_3$, a conclusão do lema decorre do fato de que $\mathbb{V}_1 \subseteq \mathbb{V}_1^*$ e $\mathbb{V}_2 \subseteq \mathbb{V}_2^*$. \square

Teorema A.3 (?[p.66]). *Considere que (\mathcal{G}, f_β) é um CM linear Gaussiano ([Definição 2.25](#)) de parâmetros $\sigma^2 = 1$, β e $\mu = 0$. Para qualquer distribuição contínua sobre β , tem probabilidade 0 a ocorrência de valores de β tais que existam $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ com $\mathbf{X} \perp^{f_\beta} \mathbf{Y} | \mathbf{Z}$ mas \mathbf{X} e \mathbf{Y} não serem d -separados dado \mathbf{Z} em \mathcal{G} .*

Lema A.4. *Se \mathbb{V}_1 não é d -separado de \mathbb{V}_2 dado \mathbb{V}_3 segundo o DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, então existe f compatível com \mathcal{G} tal que \mathbb{V}_1 e \mathbb{V}_2 são condicionalmente dependentes dado \mathbb{V}_3 segundo f*

Demonstração. Decorre do [Teorema A.3](#). \square

Prova do [Teorema 2.49](#). Decorre dos [Lemas A.2](#) e [A.4](#). \square

A.4. Relativas à [Seção 3.1](#) (O modelo de probabilidade para intervenções)

Prova do [Lema 3.2](#). Decorre da [Definição 3.1](#) que $f^*(\mathcal{V}) = \mathbb{I}(\mathbf{X} = \mathbf{x}) \prod_{V \notin \mathbf{X}} f(V | Pa(V))$. Definindo $g_{X_i}(X_i) = \mathbb{I}(X_i = x_i)$, para todo $X_i \in \mathbf{X}$ e $g_V(V, Pa(V)) = f(V | Pa(V))$, note que

$$f^*(\mathcal{V}) = \prod_{X_i \in \mathbf{X}} g_{X_i}(X_i) \prod_{V \notin \mathbf{X}} g_V(V, Pa(V))$$

Portanto, decorre do [Lema 2.14](#) que f^* é compatível com um grafo em que todo $X_i \in \mathbf{X}$ não tem pais e todo $V \notin \mathbf{X}$ tem os mesmos pais que em \mathcal{G} . Isto é, \mathcal{G} é compatível com $\mathcal{G}(\bar{\mathbf{X}})$.

Além disso, tomando $\mathbb{V} = \mathcal{V} - \mathbf{X}$,

$$\begin{aligned} f^*(\mathbf{X}) &= \int f^*(sV) d\mathbb{V} \\ &= \int \mathbb{I}(\mathbf{X} = \mathbf{x}) \prod_{V \in \mathbb{V}} f(V | Pa(V)) d\mathbb{V} \\ &= \mathbb{I}(\mathbf{X} = \mathbf{x}) \int \prod_{V \in \mathbb{V}} f(V | Pa(V)) d\mathbb{V} \\ &= \mathbb{I}(\mathbf{X} = \mathbf{x}). \end{aligned}$$

Portanto, \mathbf{X} é degenerado em \mathbf{x} segundo f^* . \square

A.4.1. Relativas ao [Teorema 3.6](#)

Lema A.5. *Considere que (\mathcal{G}, f) é um CM linear Gaussiano e que $f^* = f(\mathcal{V} | do(\mathbf{X} = \mathbf{x}))$. Se $\mathcal{G}(\bar{\mathbf{X}})$ é como no [Lema 3.2](#), então $(\mathcal{G}(\bar{\mathbf{X}}), f^*)$ é um CM linear Gaussiano tal que, para todo $V \notin \mathbf{X}$, $\mathbb{E}_f[V | Pa(V)] = \mathbb{E}_{f^*}[V | Pa(V)]$ e $\mathbb{E}[\mathbf{X}] = \mathbf{x}$.*

Demonstração. Decorre do [Lema 3.2](#) que f^* é compatível com $\mathcal{G}(\bar{\mathbf{X}})$. Além disso, também decorre do [Lema 3.2](#) que para todo $V \notin \mathbf{X}$, $f(V | Pa(V)) = f^*(V | Pa(V))$. Portanto, $\mathbb{E}_f[V | Pa(V)] = \mathbb{E}_{f^*}[V | Pa(V)]$. Finalmente, segundo o [Lema 3.2](#), \mathbf{X} é degenerado em \mathbf{x} . Assim, $\mathbb{E}[\mathbf{X}] = \mathbf{x}$. \square

Prova do Teorema 3.6. Defina $f_x^* = f(\mathcal{V}|do(X = x))$. Assim,

$$\begin{aligned} ACE_{X,Y} &= \frac{d\mathbb{E}[Y|do(X = x)]}{dx} \\ &= \frac{d\mathbb{E}_{f_x^*}[Y]}{dx} \end{aligned} \tag{A.6}$$

Além disso, decorre do [Lema A.5](#) que f^* é um CM linear Gaussiano no grafo $\mathcal{G}(\bar{X})$.

Como X não tem pais no grafo $\mathcal{G}(\bar{X})$, os únicos caminhos direcionados de X a Y que passam por X são aqueles que se iniciam em X . Formalmente, defina \mathbb{C} como o conjunto de todos os caminhos direcionados em \mathcal{G} . Além disso, $\mathbb{C}_X = \{C \in \mathbb{C} : C_i = X, \text{ para algum } i\}$. Obtemos

$$\cup_{V \in \mathcal{V}} \mathbb{C}_{V,Y} \cap \mathbb{C}_X = \mathbb{C}_{X,Y}. \tag{A.7}$$

Portanto,

$$\begin{aligned} ACE_{X,Y} &= \frac{d\mathbb{E}_{f_x^*}[Y]}{dx} && \text{likning (A.6)} \\ &= \frac{d \sum_{V \in \mathcal{V}} \sum_{C \in \mathbb{C}_{V,Y}} \mu_V \cdot \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i}}{dx} && \text{Lemas 2.27 e A.5} \\ &= \frac{d \sum_{C \in \mathbb{C}_{X,Y}} \mu_X \cdot \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i}}{dx} && \text{likning (A.7)} \\ &= \frac{d \sum_{C \in \mathbb{C}_{X,Y}} x \cdot \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i}}{dx} && \text{Lema A.5} \\ &= \sum_{C \in \mathbb{C}_{X,Y}} \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i} \end{aligned}$$

□

A.5. Seção 3.2 (Controlando confundidores (critério backdoor))

A.5.1. Teorema 3.19

Para realizar a demonstração do [Teorema 3.19](#), consideraremos um SCM aumentado, em que existe uma variável que representa a ocorrência de uma intervenção em X . Uma consequência interessante desta construção será a de que o modelo intervencional é equivalente ao condicionamento usual no SCM aumentado.

Definição A.6. Seja (\mathcal{G}_*, f_*) um SCM expandido tal que $\mathcal{G}_* = (\mathcal{V} \cup \{I_X : X \in \mathbf{X}\}, \mathcal{E}_*)$, e $\mathcal{E}_* = \mathcal{E} \cup \{(I_X \rightarrow X : X \in \mathbf{X})\}$. Isto é, \mathcal{G}_* é uma cópia de \mathcal{G} em que adicionamos para cada $X \in \mathbf{X}$ os vértice $I_X \in \{0, 1\}$ e arestas de I_X para X .

\mathcal{G}_* admite uma interpretação intuitiva. I_X é a indicadora de que fazemos uma intervenção em X , fazendo que esta assuma o valor x . Se $I_X = 0$, não há uma intervenção e, assim, X segue a sua distribuição observacional. Se $I_X = 1$, X assume o valor x com probabilidade 1.

Finalmente, considerando $Pa(X)$ como os pais de X segundo \mathcal{G} , definimos que:

$$f_*(X|Pa(X), I_X) = \begin{cases} f(X|Pa(X)) & , \text{ se } I_X = 0, \text{ e} \\ \mathbb{I}(X = x) & , \text{ caso contrário.} \end{cases}$$

Lema A.7. Se (\mathcal{G}_*, f_*) é tal qual em [Definição A.6](#), então:

$$f(\mathcal{V}|do(X = x)) = f_*(\mathcal{V}|I_{\mathbf{X}} = 1)$$

Demonstração.

$$\begin{aligned} f_*(\mathcal{V}|I_{\mathbf{X}} = 1) &= \frac{f_*(\mathcal{V}, I_{\mathbf{X}} = 1)}{f(I_{\mathbf{X}} = 1)} \\ &= \frac{f(I_{\mathbf{X}} = 1) \prod_{X \in \mathbf{X}} \mathbb{I}(X = x) \prod_{V \notin \mathbf{X}} f(V|Pa(V))}{f(I = 1)} && \text{Definição A.6} \\ &= \prod_{X \in \mathbf{X}} \mathbb{I}(X = x) \cdot \prod_{V \notin \mathbf{V}_1} f(V|Pa(V)) \\ &= f(\mathcal{V}|do(\mathbf{X} = \mathbf{x})) && \text{Definição 3.1} \end{aligned}$$

□

Lema A.8. Se (\mathcal{G}_*, f_*) é tal qual em [Definição A.6](#), então:

$$f_*(\mathcal{V}|I_{\mathbf{X}} = 0) = f(\mathcal{V}).$$

Demonstração.

$$\begin{aligned} f_*(\mathcal{V}|I_{\mathbf{X}} = 0) &= \frac{f_*(\mathcal{V}, I_{\mathbf{X}} = 0)}{f_*(I_{\mathbf{X}} = 0)} \\ &= \frac{f_*(I_{\mathbf{X}} = 0) \prod_{X \in \mathbf{X}} f_*(X|Pa(X), I_X = 0) \prod_{V \notin \mathbf{X}} f(V|Pa(V))}{f_*(I = 0)} && \text{Definição A.6} \\ &= \prod_{X \in \mathbf{X}} f(X|Pa(X)) \prod_{V \notin \mathbf{X}} f(V|Pa(V)) && \text{Definição A.6} \\ &= \prod_{V \in \mathcal{V}} f(V|Pa(V)) \\ &= f(\mathcal{V}) && \text{Definição 2.11} \end{aligned}$$

□

Lema A.9. Se (\mathcal{G}_*, f_*) é tal qual em [Definição A.6](#) e \mathbf{Z} satisfaz o segundo item do critério backdoor para medir o efeito causal de X em Y , então $I \perp^d Y|X, \mathbf{Z}$.

Demonstração. Tome um caminho arbitrário de I em Y , $C = (I, C_2, \dots, C_{n-1}, Y)$. Por definição de I , $C_2 = X$ e $I \rightarrow X$. Se $X \rightarrow C_3$, então X não é um colisor em C e C está bloqueado dado X e \mathbf{Z} . Se $X \leftarrow C_3$, então $(X, C_3, \dots, C_{n-1}, Y)$ está bloqueado dado \mathbf{Z} , uma vez que \mathbf{Z} satisfaz o segundo item do critério backdoor. Conclua que C está bloqueado dado X e \mathbf{Z} . □

Lema A.10. Se \mathbf{Z} satisfaz o segundo item do critério backdoor para medir o efeito causal de X em Y , então

$$f(y|do(x), \mathbf{z}) = f(y|x, \mathbf{z}).$$

Demonstração.

$$\begin{aligned} f(y|do(x), \mathbf{z}) &= f_*(y|I = 1, \mathbf{z}) && \text{Lema A.7} \\ &= \int f_*(y, X|I = 1, \mathbf{z})dX \\ &= \int f_*(X|I = 1, \mathbf{z})f_*(y|X, I = 1, \mathbf{z})dX \\ &= \int \mathbb{I}(X = x)f_*(y|X, I = 1, \mathbf{z})dX \\ &= f_*(y|x, I = 1, \mathbf{z}) \\ &= f_*(y|x, I = 0, \mathbf{z}) && \text{Lema A.9} \\ &= f(y|x, \mathbf{z}) && \text{Lema A.8} \end{aligned}$$

□

Lema A.11. Se (\mathcal{G}_*, f_*) é tal qual em [Definição A.6](#) e \mathbf{Z} satisfaz o critério backdoor para medir o efeito causal de X em Y , então $I \perp^d \mathbf{Z}$.

Demonstração. Tome arbitrariamente um $Z \in \mathbf{Z}$ e um caminho de I em Z , $C = (I, C_2, \dots, C_{n-1}, Z)$. Por definição de I , $C_2 = X$ e $I \rightarrow X$. Suponha por absurdo que C não tem colisor. Como, $I \rightarrow X$, decorre que $C = I \rightarrow X \rightarrow \dots \rightarrow C_{n-1} \rightarrow Z$. Assim, Z é um descendente de X , uma contradição com o critério backdoor ([Definição 3.11](#)). Conclua que C tem um colisor. Assim, C está marginalmente bloqueado ([Definição 2.47](#)). □

Lema A.12. Se \mathbf{Z} satisfaz o critério backdoor para medir o efeito causal de X em Y , então $f(\mathbf{z}|do(x)) = f(\mathbf{z})$.

Demonstração.

$$\begin{aligned} f(\mathbf{z}|do(x)) &= f_*(\mathbf{z}|I = 1) && \text{Lema A.7} \\ &= f_*(\mathbf{z}|I = 0) && \text{Lema A.11} \\ &= f(\mathbf{z}) && \text{Lema A.8} \end{aligned}$$

□

Prova do Teorema 3.19. Decorre diretamente dos [Lemas A.10](#) e [A.12](#). □

Prova do Corolário 3.20.

$$\begin{aligned} f(y|do(X = x)) &= \int f(y, \mathbf{z}|do(X = x))d\mathbf{z} \\ &= \int f(\mathbf{z}|do(X = x))f(y|do(X = x), \mathbf{z}) \\ &= \int f(\mathbf{z})f(y|x, \mathbf{z}) && \text{Teorema 3.19} \end{aligned}$$

□

A.5.2. Teoremas 3.21 e 3.22

Prova do Teorema 3.21.

$$\begin{aligned}
 \mathbb{E}[g(Y)|do(X = x), \mathbf{Z}] &= \int g(y)f(y|do(x), \mathbf{Z})dy && \text{Definição 3.3} \\
 &= \int g(y)f(y|x, \mathbf{Z})dy && \text{Teorema 3.19} \\
 &= \mathbb{E}[g(Y)|X = x, \mathbf{Z}] \\
 \mathbb{E}[g(Y)|do(X = x)] &= \mathbb{E}[\mathbb{E}[g(Y)|do(X = x), \mathbf{Z}]] \\
 &= \mathbb{E}[\mathbb{E}[g(Y)|X = x, \mathbf{Z}]] && \text{likning (A.8)}
 \end{aligned}
 \tag{A.8}$$

□

Prova do Teorema 3.22.

$$\begin{aligned}
 \mathbb{E}[g(Y)|do(x), \mathbf{Z}] &= \int g(y)f(y|do(x), \mathbf{Z})dy && \text{Definição 3.3} \\
 &= \int g(y)f(y|x, \mathbf{Z})dy && \text{Teorema 3.19} \\
 &= \int \frac{g(y)f(y, x|\mathbf{Z})}{f(x|\mathbf{Z})}dy \\
 &= \int \frac{g(y)\mathbb{I}(x_* = x)f(y, x_*|\mathbf{Z})}{f(x|\mathbf{Z})}d(x_*, y) \\
 &= \mathbb{E}\left[\frac{g(Y)\mathbb{I}(X = x)}{f(x|\mathbf{Z})}\middle|\mathbf{Z}\right] \\
 &= \frac{\mathbb{E}[g(Y)\mathbb{I}(X = x)|\mathbf{Z}]}{f(x|\mathbf{Z})}
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{E}[Y|do(x)] &= \mathbb{E}[\mathbb{E}[Y|do(X), \mathbf{Z}]] \\
 &= \mathbb{E}\left[\frac{\mathbb{E}[g(Y)\mathbb{I}(X = x)|\mathbf{Z}]}{f(x|\mathbf{Z})}\right] && \text{Teorema 3.22} \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{g(Y)\mathbb{I}(X = x)}{f(x|\mathbf{Z})}\middle|\mathbf{Z}\right]\right] \\
 &= \mathbb{E}\left[\frac{g(Y)\mathbb{I}(X = x)}{f(x|\mathbf{Z})}\right]
 \end{aligned}$$

□

A.5.3. Teorema 3.26

Lema A.13. Se $(W_n)_{n \in \mathbb{N}}$ é uma sequência de variáveis aleatórias tais que $\mathbb{E}[|W_n|] = o(1)$, então $W_n \xrightarrow{\mathbb{P}} 0$.

Demonstração.

$$\begin{aligned}
 \mathbb{P}(|W_n| > \epsilon) &\leq \frac{\mathbb{E}[|W_n|]}{\epsilon} && \text{Markov} \\
 &= o(1)
 \end{aligned}$$

□

Prova do Teorema 3.26. Como $\mathbb{E}[|\mu(x, \mathbf{Z})|] < \infty$, pela Lei dos Grandes Números,

$$\frac{\sum_{i=1}^n \mu(x, \mathbf{Z}_i)}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[\mu(x, \mathbf{Z})]$$

Portanto, pelo Teorema 3.21, é suficiente provar que $\widehat{\mathbb{E}}_1[Y|do(X=x)] - \frac{\sum_{i=1}^n \mu(x, \mathbf{Z}_i)}{n} \xrightarrow{\mathbb{P}} 0$. Usando o Lema A.13, é suficiente provar que $\mathbb{E} \left[\left| \widehat{\mathbb{E}}_1[Y|do(X=x)] - \frac{\sum_{i=1}^n \mu(x, \mathbf{Z}_i)}{n} \right| \right] = o(1)$.

$$\begin{aligned} \mathbb{E} \left[\left| \widehat{\mathbb{E}}_1[Y|do(X=x)] - \frac{\sum_{i=1}^n \mu(x, \mathbf{Z}_i)}{n} \right| \right] &= \mathbb{E} \left[\left| \frac{\sum_{i=1}^n (\hat{\mu}(x, \mathbf{Z}_i) - \mu(x, \mathbf{Z}_i))}{n} \right| \right] \\ &\leq n^{-1} \sum_{i=1}^n \mathbb{E} [|\hat{\mu}(x, \mathbf{Z}_i) - \mu(x, \mathbf{Z}_i)|] \\ &= \mathbb{E} [|\hat{\mu}(x, \mathbf{Z}) - \mu(x, \mathbf{Z})|] \\ &= o(1) \end{aligned}$$

Definição 3.23

□

A.5.4. Teorema 3.29

Prova do Teorema 3.29. Pela Lei dos Grandes números, $n^{-1} \sum_{i=1}^n \frac{Y_i \mathbb{I}(X_i=x)}{f(x|\mathbf{Z}_i)} \xrightarrow{\mathbb{P}} \mathbb{E} \left[\frac{Y \mathbb{I}(X=x)}{f(x|\mathbf{Z})} \right]$. Como pelo Teorema 3.22 temos que $\mathbb{E} \left[\frac{Y \mathbb{I}(X=x)}{f(x|\mathbf{Z})} \right] = \mathbb{E}[Y|do(X=x)]$, usando o Lema A.13 é suficiente provar que

$$\mathbb{E} \left[\left| n^{-1} \sum_{i=1}^n \frac{Y_i \mathbb{I}(X_i=x)}{\hat{f}(x|\mathbf{Z}_i)} - n^{-1} \sum_{i=1}^n \frac{Y_i \mathbb{I}(X_i=x)}{f(x|\mathbf{Z}_i)} \right| \right] = o(1).$$

$$\mathbb{E} \left[\left| n^{-1} \sum_{i=1}^n \frac{Y_i \mathbb{I}(X_i=x)}{\hat{f}(x|\mathbf{Z}_i)} - n^{-1} \sum_{i=1}^n \frac{Y_i \mathbb{I}(X_i=x)}{f(x|\mathbf{Z}_i)} \right| \right]$$

$$\leq n^{-1} \sum_{i=1}^n \mathbb{E} \left[\left| \frac{Y_i \mathbb{I}(X_i=x)}{\hat{f}(x|\mathbf{Z}_i)} - \frac{Y_i \mathbb{I}(X_i=x)}{f(x|\mathbf{Z}_i)} \right| \right]$$

$$= \mathbb{E} \left[\left| \frac{Y_1 \mathbb{I}(X_1=x)}{\hat{f}(x|\mathbf{Z}_1)} - \frac{Y_1 \mathbb{I}(X_1=x)}{f(x|\mathbf{Z}_1)} \right| \right]$$

Definição 3.23

$$= \mathbb{E} \left[\left| \frac{Y_i \mathbb{I}(X_i=x)(\hat{f}(x|\mathbf{Z}_i) - f(x|\mathbf{Z}_i))}{\hat{f}(x|\mathbf{Z}_i)f(x|\mathbf{Z}_i)} \right| \right]$$

$$\leq \delta^{-2} \mathbb{E} \left[|Y_i \mathbb{I}(X_i=x)(\hat{f}(x|\mathbf{Z}_i) - f(x|\mathbf{Z}_i))| \right]$$

 $\inf_z \min\{f(x|\mathbf{Z}_1), \hat{f}(x|\mathbf{Z}_1)\} > \delta$

$$= \delta^{-2} \mathbb{E} \left[|\hat{f}(x|\mathbf{Z}_i) - f(x|\mathbf{Z}_i)| \cdot \mathbb{E}[|Y_i \mathbb{I}(X_i=x)| | \mathbf{Z}] \right]$$

Lei da esperança total

$$\leq M \delta^{-2} \mathbb{E} \left[|\hat{f}(x|\mathbf{Z}_i) - f(x|\mathbf{Z}_i)| \right]$$

 $\sup_z \mathbb{E}[|Y_i \mathbb{I}(X_i=x)| | \mathbf{Z} = \mathbf{z}] < M$

$$= o(1)$$

□

A.5.5. Teorema 3.32

Prova do Teorema 3.32. Se as condições do [Teorema 3.26](#) estão satisfeitas, então decorre deste resultado que $\widehat{\mathbb{E}}_1[Y|do(X = x)] \xrightarrow{\mathbb{P}} \mathbb{E}[Y|do(X = x)]$. Portanto, usando [Lema A.13](#), resta demonstrar que

$$\mathbb{E} \left[\left| \widehat{\mathbb{E}}_2[Y|do(X = x)] - \sum_{i=1}^n \frac{\mathbb{I}(X_i = x)\hat{\mu}(x, \mathbf{Z}_i)}{n\hat{f}(x|\mathbf{Z}_i)} \right| \right] = o(1)$$

$$\mathbb{E} \left[\left| \widehat{\mathbb{E}}_2[Y|do(X = x)] - \sum_{i=1}^n \frac{\mathbb{I}(X_i = x)\hat{\mu}(x, \mathbf{Z}_i)}{n\hat{f}(x|\mathbf{Z}_i)} \right| \right]$$

$$= \mathbb{E} \left[\left| \sum_{i=1}^n \frac{\mathbb{I}(X_i = x)(Y_i - \hat{\mu}(x, \mathbf{Z}_i))}{n\hat{f}(x|\mathbf{Z}_i)} \right| \right]$$

Definição 3.28

$$\leq n^{-1} \sum_{i=1}^n \mathbb{E} \left[\left| \frac{\mathbb{I}(X_i = x)(Y_i - \hat{\mu}(x, \mathbf{Z}_i))}{\hat{f}(x|\mathbf{Z}_i)} \right| \right]$$

$$= \mathbb{E} \left[\left| \frac{\mathbb{I}(X_1 = x)(Y_1 - \hat{\mu}(x, \mathbf{Z}_1))}{\hat{f}(x|\mathbf{Z}_1)} \right| \right]$$

Definição 3.23

$$\leq \delta^{-1} \mathbb{E} [|\mathbb{I}(X_1 = x)(Y_1 - \hat{\mu}(x, \mathbf{Z}_1))|]$$

$\inf_{\mathbf{z}} \hat{f}(x|\mathbf{z}) > \delta$

$$\leq \delta^{-1} \mathbb{E} [|\mathbb{I}(X_1 = x)(\mathbb{E}[Y_1|X_1, \mathbf{Z}_1] - \hat{\mu}(x, \mathbf{Z}_1))|]$$

Lei da esperança total

$$= \delta^{-1} \mathbb{E} [|\mathbb{I}(X_1 = x)(\mathbb{E}[Y_1|X_1 = x, \mathbf{Z}_1] - \hat{\mu}(x, \mathbf{Z}_1))|] \quad \mathbb{I}(X_1 = x)\mathbb{E}[Y_1|X_1, \mathbf{Z}_1] \equiv \mathbb{I}(X_1 = x)\mathbb{E}[Y_1|X_1 = x, \mathbf{Z}_1]$$

$$\leq \delta^{-1} \mathbb{E} [|\mathbb{E}[Y_1|X_1 = x, \mathbf{Z}_1] - \hat{\mu}(x, \mathbf{Z}_1)|]$$

$$= \mathbb{E} [|\mu(x, \mathbf{Z}_1) - \hat{\mu}(x, \mathbf{Z}_1)|] = o(1)$$

A seguir, se as condições do [Teorema 3.29](#) estão satisfeitas, então decorre deste resultado que $\widehat{\mathbb{E}}_2[Y|do(X = x)] \xrightarrow{\mathbb{P}} \mathbb{E}[Y|do(X = x)]$. Portanto, usando [Lema A.13](#), resta demonstrar que

$$\mathbb{E} \left[\left| \widehat{\mathbb{E}}_1[Y|do(X = x)] - \sum_{i=1}^n \frac{\mathbb{I}(X_i = x)\hat{\mu}(x, \mathbf{Z}_i)}{n\hat{f}(x|\mathbf{Z}_i)} \right| \right] = o(1)$$

$$\begin{aligned}
& \mathbb{E} \left[\left| \widehat{\mathbb{E}}_1[Y|do(X=x)] - \sum_{i=1}^n \frac{\mathbb{I}(X_i=x)\hat{\mu}(x, \mathbf{Z}_i)}{n\hat{f}(x|\mathbf{Z}_i)} \right| \right] \\
&= \mathbb{E} \left[\left| \sum_{i=1}^n \frac{(\hat{f}(x|\mathbf{Z}_i) - \mathbb{I}(X_i=x))\hat{\mu}(x, \mathbf{Z}_i)}{n\hat{f}(x|\mathbf{Z}_i)} \right| \right] && \text{Definição 3.25} \\
&\leq n^{-1} \sum_{i=1}^n \mathbb{E} \left[\left| \frac{(\hat{f}(x|\mathbf{Z}_i) - \mathbb{I}(X_i=x))\hat{\mu}(x, \mathbf{Z}_i)}{\hat{f}(x|\mathbf{Z}_i)} \right| \right] \\
&= \mathbb{E} \left[\left| \frac{(\hat{f}(x|\mathbf{Z}_1) - \mathbb{I}(X_1=x))\hat{\mu}(x, \mathbf{Z}_1)}{\hat{f}(x|\mathbf{Z}_1)} \right| \right] && \text{Definição 3.23} \\
&\leq \delta^{-1} \mathbb{E} \left[|(\hat{f}(x|\mathbf{Z}_1) - \mathbb{I}(X_1=x))\hat{\mu}(x, \mathbf{Z}_1)| \right] && \inf_{\mathbf{z}} \hat{f}(x|\mathbf{Z}_1) > \delta \\
&\leq \delta^{-1} M \mathbb{E} \left[|\hat{f}(x|\mathbf{Z}_1) - \mathbb{I}(X_1=x)| \right] && \sup_{\mathbf{z}} \hat{\mu}(x, \mathbf{z}) < M \\
&= \delta^{-1} M \mathbb{E} \left[|\hat{f}(x|\mathbf{Z}_1) - \mathbb{E}[\mathbb{I}(X_1=x)|\mathbf{Z}_1]| \right] && \text{Lei da esperança total} \\
&= \delta^{-1} M \mathbb{E} \left[|\hat{f}(x|\mathbf{Z}_1) - f(x|\mathbf{Z}_1)| \right] = o(1)
\end{aligned}$$

□

A.5.6. Teorema 3.39

Prova do Teorema 3.39. Se $X \equiv \mathbb{I}(\mathbf{Z} \geq \mathbf{z}_1)$, então:

$$\begin{aligned}
CACE(\mathbf{Z} = \mathbf{z}_1) &= \mathbb{E}[Y|do(X=1), \mathbf{Z} = \mathbf{z}_1] - \mathbb{E}[Y|do(X=0), \mathbf{Z} = \mathbf{z}_1] && \text{Definição 3.7} \\
&= \lim_{\mathbf{z} \downarrow \mathbf{z}_1} \mathbb{E}[Y|do(X=1), \mathbf{Z} = \mathbf{z}] - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} \mathbb{E}[Y|do(X=0), \mathbf{Z} = \mathbf{z}] && \text{continuidade} \\
&= \lim_{\mathbf{z} \downarrow \mathbf{z}_1} \mathbb{E}[Y|X=1, \mathbf{Z} = \mathbf{z}] - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} \mathbb{E}[Y|X=0, \mathbf{Z} = \mathbf{z}] && \text{Teorema 3.21} \\
&= \lim_{\mathbf{z} \downarrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] && X \equiv \mathbb{I}(\mathbf{Z} \geq \mathbf{z}_1)
\end{aligned}$$

A seguir, considere que $f(x|\mathbf{Z}) \in (0, 1)$ é contínua exceto em \mathbf{z}_1 . Primeiramente, note que

$$\begin{aligned}
\mathbb{E}[Y|\mathbf{Z}] &= \mathbb{E}[\mathbb{E}[Y|X, \mathbf{Z}|\mathbf{Z}]] \\
&= \mathbb{E}[Y|X=1, \mathbf{Z}]f(X=1|\mathbf{Z}) + \mathbb{E}[Y|X=0, \mathbf{Z}](1 - f(X=1|\mathbf{Z})) \\
&= (\mathbb{E}[Y|X=1, \mathbf{Z}] - \mathbb{E}[Y|X=0, \mathbf{Z}])f(X=1|\mathbf{Z}) + \mathbb{E}[Y|X=0, \mathbf{Z}] \\
&= CACE(\mathbf{Z})f(X=1|\mathbf{Z}) + \mathbb{E}[Y|do(X=0), \mathbf{Z}] && \text{Teorema 3.21} \quad (\text{A.9})
\end{aligned}$$

Como $\mathbb{E}[Y|do(X=0), \mathbf{Z}]$ e $\mathbb{E}[Y|do(X=1), \mathbf{Z}]$ são contínuas em \mathbf{z}_1 , $CACE(\mathbf{Z})$ também é contínua em \mathbf{z}_1 . Assim, decorre da [likning \(A.9\)](#) que

$$\begin{aligned}
\lim_{\mathbf{z} \downarrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] &= CACE(\mathbf{z}_1) \lim_{\mathbf{z} \downarrow \mathbf{z}_1} f(X=1|\mathbf{z}) + \mathbb{E}[Y|do(X=0), \mathbf{Z} = \mathbf{z}_1] \\
\lim_{\mathbf{z} \uparrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] &= CACE(\mathbf{z}_1) \lim_{\mathbf{z} \uparrow \mathbf{z}_1} f(X=1|\mathbf{z}) + \mathbb{E}[Y|do(X=0), \mathbf{Z} = \mathbf{z}_1]
\end{aligned}$$

Finalmente subtraindo as equações acima, obtemos

$$\begin{aligned} \lim_{\mathbf{z} \downarrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] &= CACE(\mathbf{z}_1) (\lim_{\mathbf{z} \downarrow \mathbf{z}_1} f(X = 1|\mathbf{z}) - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} f(X = 1|\mathbf{z})) \\ CACE(\mathbf{z}_1) &= \frac{\lim_{\mathbf{z} \downarrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}] - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} \mathbb{E}[Y|\mathbf{Z} = \mathbf{z}]}{\lim_{\mathbf{z} \downarrow \mathbf{z}_1} f(X = 1|\mathbf{z}) - \lim_{\mathbf{z} \uparrow \mathbf{z}_1} f(X = 1|\mathbf{z})} \end{aligned}$$

□

A.6. Seções 3.3 e 3.4 (Controlando mediadores (critério frontdoor))

A.6.1. Teorema 3.48

Lema A.14. Se $\mathbf{Y} \perp^d \mathbf{Z}|\mathbf{X} \cup \mathbf{W}$ em $\mathcal{G}(\bar{\mathbf{X}})$, então

$$f(\mathbf{Y}|do(\mathbf{X}), \mathbf{Z}, \mathbf{W}) = f(\mathbf{Y}|do(\mathbf{X}), \mathbf{W})$$

Demonstração. Seja $f^*(\mathcal{V}) \equiv f(\mathcal{V}|do(\mathbf{X} = \mathbf{x}))$. Decorre do [Lema 3.2](#) que f^* é compatível com $\mathcal{G}(\bar{\mathbf{X}})$ e que \mathbf{X} é degenerado em \mathbf{x} segundo f^* . Assim,

$$\begin{aligned} f(\mathbf{Y}|do(\mathbf{X}), \mathbf{Z}, \mathbf{W}) &= f^*(\mathbf{Y}|\mathbf{Z}, \mathbf{W}) \\ &= f^*(\mathbf{Y}|\mathbf{X} = \mathbf{x}, \mathbf{Z}, \mathbf{W}) && \mathbf{X} \text{ é degenerado segundo } f^* \\ &= f^*(\mathbf{Y}|\mathbf{X} = \mathbf{x}, \mathbf{W}) && f^* \text{ compatível com } \mathcal{G}(do(\mathbf{X})), \\ & && \mathbf{Y} \perp^d \mathbf{Z}|\mathbf{X} \cup \mathbf{W} \text{ em } \mathcal{G}(do(\mathbf{X})), \text{ e } \text{Teorema 2.49} \\ &= f(\mathbf{Y}|do(\mathbf{X} = \mathbf{x}), \mathbf{W}). \end{aligned}$$

□

Lema A.15. Se $Y \perp^d \mathbf{W}|\mathbf{Z} \cup \mathbf{X}$ em $\mathcal{G}(\bar{\mathbf{X}}, \underline{\mathbf{W}})$, então

$$f(Y|do(\mathbf{X}), do(\mathbf{W}), \mathbf{Z}) = f(Y|do(\mathbf{X}), \mathbf{W}, \mathbf{Z})$$

Demonstração. Seja $f^*(\mathcal{V}) \equiv f(\mathcal{V}|do(\mathbf{X} = \mathbf{x}))$. Como $Y \perp^d \mathbf{W}|\mathbf{Z} \cup \mathbf{X}$ em $\mathcal{G}(\bar{\mathbf{X}}, \underline{\mathbf{W}})$, não há nenhum caminho ativo em $\mathcal{G}(\bar{\mathbf{X}})$ de \mathbf{X} em Y que inicia com $\mathbf{W} \leftarrow$. Isto é, $\mathbf{X} \cup \mathbf{Z}$ satisfaz o segundo item do critério backdoor para medir o efeito causal de \mathbf{W} em Y . Portanto,

$$\begin{aligned} f(Y|do(\mathbf{X} = \mathbf{x}), do(\mathbf{W}), \mathbf{Z}) &= f^*(Y|do(\mathbf{W}), \mathbf{Z}) \\ &= f^*(Y|do(\mathbf{W}), \mathbf{X} = \mathbf{x}, \mathbf{Z}) && \text{Lema 3.2} \\ &= f^*(Y|\mathbf{W}, \mathbf{X} = \mathbf{x}, \mathbf{Z}) && \text{Lema A.10} \\ &= f^*(Y|\mathbf{W}, \mathbf{Z}) \\ &= f(Y|do(\mathbf{X} = \mathbf{x}), \mathbf{W}, \mathbf{Z}) \end{aligned}$$

□

Lema A.16. Se $\mathbf{Y} \perp^d I_{\mathbf{X}} | \mathbf{Z} \cup \mathbf{W}$ em $\mathcal{G}(\bar{\mathbf{W}}, \mathbf{X}^+)$, então:

$$f(Y|do(\mathbf{W}), do(\mathbf{X}), \mathbf{Z}) = f(Y|do(\mathbf{W}), \mathbf{Z})$$

Demonstração. Seja $f^*(\mathcal{V}) \equiv f(\mathcal{V}|do(\mathbf{W} = \mathbf{w}))$.

$$\begin{aligned} f(Y|do(\mathbf{W} = \mathbf{w}), do(\mathbf{X}), \mathbf{Z}) &= f^*(Y|do(\mathbf{X}), \mathbf{Z}) && \text{Lema 3.2} \\ &= f^*(Y|do(\mathbf{X}), \mathbf{W} = \mathbf{w}, \mathbf{Z}) && \text{Lema 3.2} \\ &= f^*(Y|I_{\mathbf{X}} = 1, \mathbf{W} = \mathbf{w}, \mathbf{Z}) && \text{Lema A.7} \\ &= f^*(Y|\mathbf{W} = \mathbf{w}, \mathbf{Z}, I_{\mathbf{X}} = 0) && \mathbf{Y} \perp^d I_{\mathbf{X}} | \mathbf{W} \cup \mathbf{Z} \text{ em } \mathcal{G}(\bar{\mathbf{W}}, \mathbf{X}^+) \\ &= f^*(Y|\mathbf{W} = \mathbf{w}, \mathbf{Z}) && \text{Lema A.8} \\ &= f(Y|do(\mathbf{W} = \mathbf{w}), \mathbf{Z}) && \text{Lema 3.2} \end{aligned}$$

□

Prova do Teorema 3.48. Decorre dos [Lemas A.14](#) a [A.16](#).

□

A.6.2. Teorema 3.44

Lema A.17. Se \mathbf{W} satisfaz o critério frontdoor para medir o efeito causal de X em Y , então $f(Y|do(X), \mathbf{W}) = f(Y|do(X), do(\mathbf{W}))$.

Demonstração. Decorre do critério frontdoor [Definição 3.43.3](#) que X satisfaz o item 2 do critério backdoor para medir o efeito causal de \mathbf{W} em Y . Portanto, pelo [Lema 3.49](#), $\mathbf{Y} \perp^d \mathbf{W} | \mathbf{X}$ em $\mathcal{G}(\bar{\mathbf{W}})$. Pelo [Exercício 2.56](#), $\mathbf{Y} \perp^d \mathbf{W} | \mathbf{X}$ em $\mathcal{G}(\bar{\mathbf{X}}, \bar{\mathbf{W}})$. A prova se conclui aplicando o item 2 do [Teorema 3.48](#). □

Lema A.18. Se \mathbf{W} satisfaz o critério frontdoor para medir o efeito causal de X em Y , então $f(Y|do(X), do(\mathbf{W})) = f(Y|do(\mathbf{W}))$.

Demonstração. A prova consiste em aplicar o item 3 do [Teorema 3.48](#). Para tal, desejamos provar que $\mathbf{Y} \perp^d I_{\mathbf{X}} | \mathbf{W}$ em $\mathcal{G}(\bar{\mathbf{W}}, X^+)$. Tome C como um caminho arbitrário em $\mathcal{G}(\bar{\mathbf{W}}, X^+)$ de $I_{\mathbf{X}}$ em Y .

Primeiramente, provaremos que C não é um caminho direcionado. C não é um caminho direcionado de Y a $I_{\mathbf{X}}$ pois a única aresta em $\mathcal{G}(\bar{\mathbf{W}}, X^+)$ ligada a $I_{\mathbf{X}}$ é $I_{\mathbf{X}} \rightarrow X$. A seguir, suponha que C é um caminho direcionado de $I_{\mathbf{X}}$ a Y . Pelo [Definição 3.43.2](#), existe C_i tal que $C_i \in \mathbf{W}$. Como C é direcionado de $I_{\mathbf{X}}$ em Y , $C_{i-1} \rightarrow C_i$. Este é um absurdo, pois não há aresta apontando para \mathbf{W} em $\mathcal{G}(\bar{\mathbf{W}}, X^+)$. Portanto, C não é um caminho direcionado.

Conclua que existe C_i que é um colisor. Como não há arestas apontando para \mathbf{W} em $\mathcal{G}(\bar{\mathbf{W}}, X^+)$, não há $W \in \mathbf{W}$ que é descendente de C_i . Como C_i é um colisor e C_i não tem descendentes em \mathbf{W} , conclua que C está bloqueado. Como C era arbitrário, $\mathbf{Y} \perp^d I_{\mathbf{X}} | \mathbf{W}$ em $\mathcal{G}(\bar{\mathbf{W}}, X^+)$. □

Prova do Teorema 3.44.

$$\begin{aligned}
f(Y|do(X = x)) &= \int f(Y|do(X = x), \mathbf{W})f(\mathbf{W}|do(X = x))d\mathbf{W} \\
&= \int f(Y|do(X = x), \mathbf{W})f(\mathbf{W}|X = x)d\mathbf{W} && \text{Exercício 3.52} \\
&= \int f(Y|do(X = x), do(\mathbf{W}))f(\mathbf{W}|X = x)d\mathbf{W} && \text{Lema A.17} \\
&= \int f(Y|do(\mathbf{W}))f(\mathbf{W}|X = x)d\mathbf{W} && \text{Lema A.18} \\
&= \int \int f(Y|\mathbf{W}, X)f(X)dXf(\mathbf{W}|X = x)d\mathbf{W} && \text{Exercício 3.52}
\end{aligned}$$

□

A.6.3. Teorema 3.45

Prova do Teorema 3.45.

$$\begin{aligned}
\mathbb{E}[Y|do(X = x)] &= \int Y f(Y|do(X = x))dY \\
&= \int Y \int f(\mathbf{W}|x) \int f(Y|X, \mathbf{W})f(X)dXdWdY && \text{Teorema 3.44} \\
&= \int Y f(Y|X, \mathbf{W})f(X)f(\mathbf{W}|x)d(Y \times \mathbf{W} \times Y) \\
&= \int \frac{Y f(\mathbf{W}|x)}{f(\mathbf{W}|X)} f(Y, \mathbf{W}, X)d(Y \times \mathbf{W} \times Y) \\
&= \mathbb{E} \left[\frac{Y f(\mathbf{W}|x)}{f(\mathbf{W}|X)} \right]
\end{aligned}$$

□

A.7. Seção 4.1 (Levando a intuição do SCM ao POM)

Prova do Lema 4.10. Se $\mathbf{Z}_{\mathbb{V}=\mathbf{v}}(\omega) \neq \mathbf{z}$, então $\mathbb{I}(\mathbf{Z}_{\mathbb{V}=\mathbf{v}}(\omega) = \mathbf{z}) = 0$ e

$$W_{\mathbb{V}=\mathbf{v}}(\omega)\mathbb{I}(\mathbf{Z}_{\mathbb{V}=\mathbf{v}}(\omega) = \mathbf{z}) = 0 = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)\mathbb{I}(\mathbf{Z}_{\mathbb{V}=\mathbf{v}}(\omega) = \mathbf{z})$$

Se $\mathbf{Z}_{\mathbb{V}=\mathbf{v}}(\omega) = \mathbf{z}$, então $\mathbb{I}(\mathbf{Z}_{\mathbb{V}=\mathbf{v}}(\omega) = \mathbf{z}) = 1$ e basta provar que $W_{\mathbb{V}=\mathbf{v}}(\omega) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)$.

Para tal, considere primeiramente que $W \in \mathbf{Z}$. Pela definição de ω , $\mathbf{Z}_{\mathbb{V}=\mathbf{v}}(\omega) = \mathbf{z} = \mathbf{Z}_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)$. Portanto, $W_{\mathbb{V}=\mathbf{v}}(\omega) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)$. Similarmente, se $W \in \mathbb{V}$, decorre da Definição 4.7 que $\mathbb{V}_{\mathbb{V}=\mathbf{v}}(\omega) = \mathbf{v} = \mathbb{V}_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)$. Portanto, $W_{\mathbb{V}=\mathbf{v}}(\omega) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)$. Assim, resta considerar o caso em que $W \notin (\mathbf{Z} \cup \mathbb{V})$.

Para provar este fato, construiremos uma ordem sobre $\mathcal{V} - (\mathbf{Z} \cup \mathbb{V})$. Defina $\mathcal{V}^{(0)} = \{W \in \mathcal{V} - (\mathbf{Z} \cup \mathbb{V}) : Pa(W) = \emptyset\}$, isto é $\mathcal{V}^{(0)}$ são vértices no DAG que não estão em $\mathbf{Z} \cup \mathbb{V}$ e que são raízes. Além disso, para todo $1 \leq i \leq n$, $\mathcal{V}^{(i)} = \{W \in \mathcal{V} - (\mathbf{Z} \cup \mathbb{V}) : Pa(W) \subseteq \mathcal{V}^{(i-1)} \cup (\mathbf{Z} \cup \mathbb{V})\}$. Isto é, todos os pais de $\mathcal{V}^{(1)}$ são raízes ou estão em $(\mathbf{Z} \cup \mathbb{V})$, todos os avós de $\mathcal{V}^{(2)}$ são raízes ou estão em $(\mathbf{Z} \cup \mathbb{V})$, e assim por diante. Como \mathcal{V} é finito, existe n tal que $\mathcal{V}^{(n)} \equiv \mathcal{V} - (\mathbf{Z} \cup \mathbb{V})$.

Completaremos a prova por indução finita. Primeiramente, se $W \in \mathcal{V}^{(0)}$, decorre da Definição 4.7 que $W_{\mathbb{V}=\mathbf{v}} \equiv g_W(U_W) \equiv W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}$. Em particular, $W_{\mathbb{V}=\mathbf{v}}(\omega) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)$. A seguir, suponha que para todo $W \in \mathcal{V}^{(i-1)}$,

$W_{\mathbb{V}=\mathbf{v}}(\omega) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)$ e tome $W \in \mathcal{V}^{(i)}$. Por definição de $\mathcal{V}^{(i)}$, $Pa(W) \subseteq \mathcal{V}^{(i-1)} \cup (\mathbf{Z} \cup \mathbb{V})$. Tome $W^* \in Pa(W)$. Por hipótese de indução, se $W^* \in \mathcal{V}^{(i-1)}$, então $W_{\mathbb{V}=\mathbf{v}}^*(\omega) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}^*(\omega)$. Também provamos que, se $W^* \in \mathbf{Z} \cup \mathbb{V}$, então $W_{\mathbb{V}=\mathbf{v}}^*(\omega) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}^*(\omega)$. Conclua que $Pa^*(W_{\mathbb{V}=\mathbf{v}})(\omega) = Pa^*(W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}})(\omega)$. Como decorre da [Definição 4.7](#) que $W_{\mathbb{V}=\mathbf{v}} \equiv g_W(U_W, Pa^*(W_{\mathbb{V}=\mathbf{v}}))$ e $W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}} \equiv g_W(U_W, Pa^*(W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}))$ e $Pa^*(W_{\mathbb{V}=\mathbf{v}})(\omega) = Pa^*(W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}})(\omega)$, conclua que $W_{\mathbb{V}=\mathbf{v}}(\omega) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}(\omega)$. A prova está completa observando que $W \in \mathcal{V}^{(n)} = \mathcal{V} - (\mathbf{Z} \cup \mathbb{V})$. \square

Prova do [Lema 4.9](#).

$$\begin{aligned} & \mathbb{P}(\mathcal{V}_{\mathbb{V}=\mathbf{v}} = \mathcal{V}_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}} | \mathbf{Z}_{\mathbb{V}=\mathbf{v}} = \mathbf{z}) \\ &= \mathbb{P}(W_{\mathbb{V}=\mathbf{v}} = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}}, \forall W \in \mathcal{V} | \mathbf{Z}_{\mathbb{V}=\mathbf{v}} = \mathbf{z}) \\ &= \mathbb{P}(W_{\mathbb{V}=\mathbf{v}} \mathbb{I}(\mathbf{Z}_{\mathbb{V}=\mathbf{v}} = \mathbf{z}) = W_{\mathbf{Z}=\mathbf{z}, \mathbb{V}=\mathbf{v}} \mathbb{I}(\mathbf{Z}_{\mathbb{V}=\mathbf{v}} = \mathbf{z}), \forall W \in \mathcal{V} | \mathbf{Z}_{\mathbb{V}=\mathbf{v}} = \mathbf{z}) \\ &= 1 \end{aligned}$$

[Lema 4.10](#)

\square

Prova do [Lema 4.11](#).

$$\begin{aligned} f(\mathcal{V}_{\mathbf{v}}) &= \mathbb{I}(\mathbb{V} = \mathbf{v}) \cdot \prod_{V \in \mathcal{V} - \mathbb{V}} f^*(V | Pa(V)) && \text{Definição 4.7} \\ &= \mathbb{I}(\mathbb{V} = \mathbf{v}) \cdot \prod_{V \in \mathcal{V} - \mathbb{V}} f(V | Pa(V)) && \text{Definições 4.3 e 4.7} \\ &= f(\mathcal{V} | do(\mathbb{V} = \mathbf{v})) && \text{Definição 3.1} \end{aligned}$$

\square

Prova do [Lema 4.13](#). (1 \rightarrow 2) Para realizar esta demonstração provaremos que a negação de 2 implica a negação de 1. Suponha que exista um ascendente comum de X e Y . Portanto, existe $V \in \mathcal{V}$, um caminho direcionado de V em X , $C^X = (V, C_2^X, \dots, C_{n-1}^X, X)$, e um caminho direcionado de V em Y , $C^Y = (V, C_2^Y, \dots, C_{m-1}^Y, Y)$. Defina $C = (X, C_{n-1}^X, \dots, C_2^X, V, C_2^Y, \dots, C_{m-1}^Y, Y)$. Como C^X é um caminho direcionado, $(C_{n-1}^X, X) \in \mathcal{E}$. Além disso, como C^X e C^Y são caminhos direcionados, não há colisor em C . Conclua que C está bloqueado dado \emptyset . Isto é, \emptyset não satisfaz o critério backdoor para medir o efeito causal de X em Y .

(2 \rightarrow 3) Para realizar esta demonstração provaremos que a negação de 3 implica a negação de 2. Suponha que no grafo potencial, \mathcal{G}^* , existe um caminho não bloqueado de X a Y_x , C . Portanto, existe $V \in \mathcal{V}$ tal que $C = X, \dots, V \leftarrow U_V \rightarrow V_x, \dots, Y_x$. Como C não está bloqueado, não há colisor em C . Portanto,

$$C = X \leftarrow \dots \leftarrow V \leftarrow U_V \rightarrow V_x \rightarrow \dots \rightarrow Y_x.$$

Decorre da [Definição 4.5](#) que V é ancestral comum de X e Y . Para constatar essa última afirmação basta remover U_V e V_x do caminho e substituir cada vértice potencial a partir de V_x por sua cópia em \mathcal{V} .

(3 \rightarrow 1) Esta demonstração decorre do [Lema 4.16](#), provado a seguir, tomando $\mathbf{Z} = \emptyset$. \square

Prova do [Lema 4.16](#). Suponha que \mathbf{Z} não satisfaz o critério backdoor para medir o efeito causal de X em Y . Como $X \notin Anc(\mathbf{Z})$, existe um caminho não bloqueado de X em Y dado \mathbf{Z} , $C = (X, C_2, \dots, C_{n-1}, Y)$ tal que $X \leftarrow C_2$. Há dois casos para considerar: $C_{n-1} \leftarrow Y$ e $C_{n-1} \rightarrow Y$.

Se $C_{n-1} \leftarrow Y$, então não há colisor em $C_{n-1} \leftarrow Y \leftarrow U_Y \rightarrow Y_x$. Portanto, $C^* = (X, C_2, \dots, C_{n-1}, Y, U_Y, Y_x)$ é um caminho desbloqueado de X a Y_x no grafo potencial.

A seguir, considere que $C_{n-1} \rightarrow Y$. Tome $m = \max(\{1\} \cup \{i : C_i \text{ é colisor}\})$. Assim, $C_m \leftarrow C_{m+1} \dots C_{n-1} \rightarrow Y$. Pelo diagrama acima, existe $p > m$ tal que $C_{p-1} \leftarrow C_p \rightarrow C_{p+1}$. Defina $C^* = (X_x, (C_2)_x, \dots, (C_{n-1})_x, Y_x)$. Não há colisor em $C_{p-1} \leftarrow C_p \leftarrow U_{C_p} \rightarrow C_p^*$. Também, como não há colisor em (C_p, C_{p+1}, \dots, Y) , decorre da [Definição 4.5](#) que não há colisor em $(C_p^*, C_{p+1}^*, \dots, Y)$. Portanto, não há colisor em $(C_p, U_{C_p}, C_p^*, C_{p+1}^*, \dots, Y)$. Defina

$$C^+ = (X, C_2, \dots, C_p, U_{C_p}, C_p^*, C_{p+1}^*, \dots, Y).$$

Como C está bloqueado dado \mathbf{Z} , para todo $i \leq p$, $C_i \in \mathbf{Z}$ se e somente se C_i é um colisor em C . Além disso, para todo $i > p$, C_i^+ não é colisor e C_i^+ não está em \mathbf{Z} , pois é um resultado potencial ou uma variável em U . Assim C_i^+ não está bloqueado dado \mathbf{Z} e é um caminho de X a Y_x . \square

A.8. Seção 4.2 (Variáveis Instrumentais)

Definição A.19. $\mathbb{U} = \{U_V : V \in \mathcal{V}\}$, $\mathcal{V}^{(1)} = \{V \in \mathcal{V} : Pa(V) = \emptyset\}$, e $\mathcal{V}^{(i)} = \{V \in \mathcal{V} : Pa(V) \subseteq \mathcal{V}^{(i-1)}\}$.

Lema A.20. No modelo de resultados potenciais ([Definição 4.7](#)) se $Anc^*(Y_{\mathbf{X}=\mathbf{x}}) \cap \mathbb{U} = Anc^*(Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) \cap \mathbb{U}$, então para cada $V \in Pa(Y)$, $Anc^*(V_{\mathbf{X}=\mathbf{x}}) \cap \mathbb{U} = Anc^*(V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) \cap \mathbb{U}$.

Demonstração. Primeiramente, note que para todo $Z \in \mathbf{Z}$, tem-se que $U_Z \notin Anc^*(Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) = Anc^*(Y_{\mathbf{X}=\mathbf{x}})$. Assim, como $Anc^*(V_{\mathbf{X}=\mathbf{x}}) \subseteq Anc^*(Y_{\mathbf{X}=\mathbf{x}})$, $U_Z \notin Anc^*(V_{\mathbf{X}=\mathbf{x}})$. Isto é, $\mathbf{Z}_{\mathbf{X}=\mathbf{x}} \cap Anc^*(V_{\mathbf{X}=\mathbf{x}}) = \emptyset$.

A seguir, por construção da [Definição 4.7](#), $Anc^*(V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) \cap \mathbb{U} \subseteq Anc^*(V_{\mathbf{X}=\mathbf{x}}) \cap \mathbb{U}$. Assim, basta provar que para todo $U \in Anc^*(V_{\mathbf{X}=\mathbf{x}}) \cap \mathbb{U}$ tem-se que $U \in Anc^*(V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}})$.

Tome $U \in Anc^*(V_{\mathbf{X}=\mathbf{x}}) \cap \mathbb{U}$. Por construção, existem vértices, $C_1, \dots, C_m \in \mathcal{V}$ que constituem um caminho direcionado de U a $V_{\mathbf{X}=\mathbf{x}}$, $(U, (C_1)_{\mathbf{X}=\mathbf{x}}, \dots, (C_m)_{\mathbf{X}=\mathbf{x}}, V_{\mathbf{X}=\mathbf{x}})$. Como $\mathbf{Z}_{\mathbf{X}=\mathbf{x}} \cap Anc^*(V_{\mathbf{X}=\mathbf{x}}) = \emptyset$, não existe V_i tal que $V_i \in \mathbf{Z}$. Portanto, $(U, (C_1)_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}, \dots, (C_m)_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}, V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}})$ é um caminho direcionado de U a $V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}$, isto é, $U \in Anc^*(V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}})$. \square

Lema A.21. No modelo de resultados potenciais ([Definição 4.7](#)), se $Anc^*(Y_{\mathbf{X}=\mathbf{x}}) \cap \mathbb{U} = Anc^*(Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) \cap \mathbb{U}$, então $Y_{\mathbf{X}=\mathbf{x}} \equiv Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}$.

Demonstração. Faremos a demonstração por indução. Para tal, utilizaremos a [Definição A.19](#). Se $Y \in \mathcal{V}^{(1)}$, então $Pa(Y) = \emptyset$. Assim, $Anc^*(Y_{\mathbf{X}=\mathbf{x}})$ é $\{U_Y\}$ ou \emptyset . Se $Anc^*(Y_{\mathbf{X}=\mathbf{x}}) = \emptyset$, então $Y \in \mathbf{X}$. Portanto, tomando Y como X_i , $Y_{\mathbf{X}=\mathbf{x}} \equiv \mathbf{x}_i \equiv Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}$. Se $Anc^*(Y_{\mathbf{X}=\mathbf{x}}) = Anc^*(Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) = \{U_Y\}$, então $Y_{\mathbf{X}=\mathbf{x}} \equiv g_Y(U_Y) \equiv Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}$.

Agora, suponha que se $V \in \mathcal{V}^{(i-1)}$ e $Anc^*(V_{\mathbf{X}=\mathbf{x}}) \cap \mathbb{U} = Anc^*(V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) \cap \mathbb{U}$, então $V_{\mathbf{X}=\mathbf{x}} \equiv V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}$. Tome $Y \in \mathcal{V}^{(i)}$ tal que $Anc^*(Y_{\mathbf{X}=\mathbf{x}}) \cap \mathbb{U} = Anc^*(Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) \cap \mathbb{U}$. Se $U_Y \notin Anc^*(Y_{\mathbf{X}=\mathbf{x}})$, então existe Y é algum X_i . Portanto, $Y_{\mathbf{X}=\mathbf{x}} \equiv \mathbf{x}_i \equiv Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}$. A seguir, suponha que $U_Y \in Anc^*(Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}})$. Para cada $V \in Pa(Y)$, como $Y \in \mathcal{V}^{(i)}$, $V \in \mathcal{V}^{(i-1)}$. Além disso, decorre do [Lema A.20](#), que $Anc^*(V_{\mathbf{X}=\mathbf{x}}) \cap U = Anc^*(V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) \cap U$. Portanto, decorre da hipótese de indução que $V_{\mathbf{X}=\mathbf{x}} \equiv V_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}$. Assim,

$$Y_{\mathbf{X}=\mathbf{x}} \equiv g_Y(U_Y, (Pa(Y))_{\mathbf{X}=\mathbf{x}}) \equiv g_Y(U_Y, (Pa(Y))_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}) \equiv Y_{\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}}$$

\square

Prova do Lema 4.23. A seguir, suponha que existe um caminho direcionado de I a Y , (I, C_1, \dots, C_m, Y) , que não passa por X . Escolha f tal que $I \sim \text{Bernoulli}(0.5)$ e $I \equiv C_1 \equiv \dots \equiv C_m \equiv Y$. $Y_{X=x} \sim \text{Bernoulli}(0.5)$ e $Y_{X=x, I=i} \equiv i$. Portanto, $\mathbb{P}(Y_{X=x} \neq Y_{X=x, I=i}) > 0$.

A seguir, suponha que todo caminho direcionado de I a Y , C , é tal que existe j com $C_j = X$. Iremos provar que $\text{Anc}^*(Y_{X=x}) \cap \mathbb{U} = \text{Anc}^*(Y_{X=x, I=i}) \cap \mathbb{U}$ e, com base no [Lema A.21](#), concluir que $Y_{I=i, X=x} \equiv Y_{X=x}$. Como $\text{Anc}^*(Y_{X=x, I=i}) \cap \mathbb{U} \subseteq \text{Anc}^*(Y_{X=x}) \cap \mathbb{U}$, basta provar que todo $U \in \text{Anc}^*(Y_{X=x}) \cap \mathbb{U}$ satisfaz $U \in \text{Anc}^*(Y_{X=x, I=i})$.

Tome $U \in \text{Anc}^*(Y_{X=x}) \cap \mathbb{U}$. Assim, existem vértices $C_1, \dots, C_m \in \mathcal{V}$ e um caminho direcionado de U a $Y_{\mathbf{X}=\mathbf{x}}$, $(U, (C_1)_{X=x}, \dots, (C_m)_{X=x}, Y_{X=x})$. Note que se algum C_j fosse I , então pela hipótese do lema, existiria algum C_k que seria X . Assim, $(U, C_1, \dots, C_m, Y_{X=x})$ não seria um caminho direcionado, afinal, $X_{X=x}$ não tem pais. Portanto, I não está em C_1, \dots, C_m . Conclua que $(U, (C_1)_{X=x, I=i}, \dots, (C_m)_{X=x, I=i}, Y_{X=x, I=i})$ é um caminho direcionado de U a $Y_{X=x, I=i}$. Isto é, $U \in \text{Anc}^*(Y_{X=x, I=i})$. Decorre do [Lema A.21](#) que $Y_{I=i, X=x} \equiv Y_{X=x}$. \square

Lema A.22. *Se I é um instrumento para medir o efeito causal de X em Y e $X \in \text{Anc}(Y)$, então I é ignorável para o efeito em X .*

Demonstração. Provaremos a contra-positiva. Se I não é ignorável para medir o efeito em X , então decorre do [Lema 4.13](#) que I e X tem um ancestral comum, Z . Como X é um ancestral de Y , decorre que Z é ancestral comum a I e Y . Portanto, conclui-se do [Lema 4.13](#) que I não é ignorável para Y . Isto é, pela [Definição 4.22.1](#), I não é um instrumento. \square

Lema A.23. *Se X é ignorável para medir o efeito causal em Y em um CM linear Gaussiano, então*

$$ACE = \text{Cov}[X, Y] \cdot \mathbb{V}^{-1}[X].$$

Demonstração. Decorre do [Lema 2.26](#) que \mathcal{V} segue uma normal multivariada. Portanto, existem α e β tais que

$$\mathbb{E}[Y|X] = \alpha + \beta \cdot X. \tag{A.10}$$

Assim,

$$\begin{aligned} ACE &= \frac{d\mathbb{E}[Y|do(X=x)]}{dx} \\ &= \frac{d\mathbb{E}[Y|X=x]}{dx} && \text{Corolário 4.14} \\ &= \frac{d(\alpha + \beta x)}{dx} = \beta && \text{likning (A.10)} \end{aligned} \tag{A.11}$$

Finalmente,

$$\begin{aligned} \text{Cov}[X, Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X\mathbb{E}[Y|X]] - \mathbb{E}[X]\mathbb{E}[\mathbb{E}[Y|X]] \\ &= \mathbb{E}[X(\alpha + \beta X)] - \mathbb{E}[X]\mathbb{E}[\alpha + \beta X] \\ &= \alpha\mathbb{E}[X] + \beta\mathbb{E}[X^2] - \alpha\mathbb{E}[X] - \beta\mathbb{E}[X]^2 \\ &= \beta\mathbb{V}[X] \\ &= ACE \cdot \mathbb{V}[X] && \text{likning (A.11)} \end{aligned}$$

Rearranjando os termos, obtenha $ACE = \text{Cov}[X, Y] \cdot \mathbb{V}^{-1}[X]$. \square

Prova do Teorema 4.24.

$$\begin{aligned}
Cov[I, Y] \cdot \mathbb{V}^{-1}[I] &= ACE_{I,Y} && \text{Lema A.23} \\
&= \sum_{C \in \mathcal{C}_{I,Y}} \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i} && \text{Teorema 3.6} \\
&= \sum_{C \in \mathcal{C}_{I,X}} \sum_{K \in \mathcal{C}_{X,Y}} \left(\prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i} \right) \left(\prod_{j=1}^{|K|-1} \beta_{K_{j+1}, K_j} \right) && \text{Lema 4.23} \\
&= \left(\sum_{C \in \mathcal{C}_{I,X}} \prod_{i=1}^{|C|-1} \beta_{C_{i+1}, C_i} \right) \left(\sum_{K \in \mathcal{C}_{X,Y}} \prod_{j=1}^{|K|-1} \beta_{K_{j+1}, K_j} \right) \\
&= ACE_{I,X} \cdot ACE_{X,Y} && \text{Teorema 3.6} \\
&= Cov[I, X] \cdot \mathbb{V}^{-1}[I] \cdot ACE_{X,Y} && \text{Lemas A.22 e A.23}
\end{aligned}$$

Rearranjando os termos, obtemos $ACE_{X,Y} = \frac{Cov[I,Y]}{Cov[I,X]}$. □

Prova do Teorema 4.27.

$$\begin{aligned}
&Y_{I=1} - Y_{I=0} \\
&= Y_{I=1} \mathbb{I}(X_{I=1} = 1) + Y_{I=1} \mathbb{I}(X_{I=1} = 0) - Y_{I=0} \mathbb{I}(X_{I=0} = 1) - Y_{I=0} \mathbb{I}(X_{I=0} = 0) \\
&= Y_{I=1, X=1} \mathbb{I}(X_{I=1} = 1) + Y_{I=1, X=0} \mathbb{I}(X_{I=1} = 0) - Y_{I=0, X=1} \mathbb{I}(X_{I=0} = 1) - Y_{I=0, X=0} \mathbb{I}(X_{I=0} = 0) && \text{Lema 4.10} \\
&= Y_{X=1} \mathbb{I}(X_{I=1} = 1) + Y_{X=0} \mathbb{I}(X_{I=1} = 0) - Y_{X=1} \mathbb{I}(X_{I=0} = 1) - Y_{X=0} \mathbb{I}(X_{I=0} = 0) && \text{Definição 4.22.2} \\
&= (Y_{X=1} - Y_{X=0}) (\mathbb{I}(X_{I=1} = 1) - \mathbb{I}(X_{I=0} = 1)) \\
&= (Y_{X=1} - Y_{X=0}) (X_{I=1} - X_{I=0}) && X \in \{0, 1\} \\
& && \text{(A.12)}
\end{aligned}$$

Portanto,

$$\begin{aligned}
\mathbb{E}[Y_{I=1} - Y_{I=0}] &= \mathbb{E}[(Y_{X=1} - Y_{X=0})(X_{I=1} - X_{I=0})] && \text{likning (A.12)} \\
&= \mathbb{E}[Y_{X=1} - Y_{X=0} | X_{I=1} - X_{I=0} = 1] \mathbb{P}(X_{I=1} - X_{I=0} = 1) && \text{Definição 4.25} \quad \text{(A.13)} \\
& && \text{(A.14)}
\end{aligned}$$

Como I é um instrumento, decorre da Definição 4.22 que $Cov[I, X] \neq 0$. Portanto, $\mathbb{P}(X_{I=1} - X_{I=0} = 1) \neq 0$. Reagrupando os termos na likning (A.13), obtemos:

$$\begin{aligned}
\mathbb{E}[Y_{X=1} - Y_{X=0} | X_{I=1} - X_{I=0} = 1] &= \frac{\mathbb{E}[Y_{I=1} - Y_{I=0}]}{\mathbb{P}(X_{I=1} - X_{I=0} = 1)} \\
& \text{LATE} = \frac{\mathbb{E}[Y_{I=1} - Y_{I=0}]}{\mathbb{P}(X_{I=1} - X_{I=0} = 1)} && \text{Definição 4.26} \quad \text{(A.15)}
\end{aligned}$$

$$= \frac{\mathbb{E}[Y_{I=1} - Y_{I=0}]}{\mathbb{E}[X_{I=1} - X_{I=0}]} \quad X \in \{0, 1\}, \text{Definição 4.25} \quad \text{(A.16)}$$

$$\text{(A.17)}$$

Há dois casos a considerar. Se $X \in \text{Anc}(Y)$. Assim, decorre do [Lema A.22](#) que I é ignorável para medir X . Neste caso, podemos continuar a desenvolver [likning \(A.15\)](#):

$$LATE = \frac{\mathbb{E}[Y|I=1] - \mathbb{E}[Y|I=0]}{\mathbb{E}[X|I=1] - \mathbb{E}[X|I=0]} \quad \text{Definição 4.22.1, Corolário 4.14}$$

Se $X \notin \text{Anc}(Y)$, então como I é um instrumento, decorre do [Lema 4.23](#) que $I \notin \text{Anc}(Y)$. Portanto, conclua do [Exercício 3.36](#) que $\mathbb{E}[Y|do(I=1)] - \mathbb{E}[Y|do(I=0)] = 0$, o que implica pelo [Lema 4.11](#) que $\mathbb{E}[Y_{I=1} - Y_{I=0}] = 0$. Como I é ignorável para Y , decorre do [Corolário 4.14](#) que $\mathbb{E}[Y|I=1] - \mathbb{E}[Y|I=0] = 0$. Assim, decorre do [likning \(A.15\)](#) que $LATE = 0 = \frac{\mathbb{E}[Y|I=1] - \mathbb{E}[Y|I=0]}{\mathbb{E}[X|I=1] - \mathbb{E}[X|I=0]}$. \square

A.9. Seção 4.3 (Contrafactuais)

Prova do Teorema 4.30. Tome um caminho arbitrário de $Y_{\mathbf{X}=\mathbf{x}}$ a \mathbf{Z} . Decorre da [Definição 4.7](#) que o caminho necessariamente passará por $U \in \mathbb{U}$. Como U é uma raiz, ele não é um colisor no caminho. Portanto, o caminho está bloqueado dado \mathbb{U} . Como o caminho era arbitrário, $Y_{\mathbf{X}=\mathbf{x}} \perp^d \mathbf{Z}|\mathbb{U}$. Decorre do [Teorema 2.49](#) que $Y_{\mathbf{X}=\mathbf{x}}$ é independente de \mathbf{Z} dado \mathbb{U} . Assim,

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{\mathbf{X}=\mathbf{x}} \leq \mathbf{y}|\mathbf{Z} = \mathbf{z}) &= \int \mathbb{P}(\mathbf{Y}_{\mathbf{X}=\mathbf{x}} \leq \mathbf{y}|\mathbf{Z} = \mathbf{z}, \mathbb{U})f(\mathbb{U}|\mathbf{Z} = \mathbf{z})d\mathbb{U} \\ &= \int \mathbb{P}(\mathbf{Y}_{\mathbf{X}=\mathbf{x}} \leq \mathbf{y}|\mathbb{U})f(\mathbb{U}|\mathbf{Z} = \mathbf{z})d\mathbb{U} \end{aligned} \quad \mathbf{Y}_{\mathbf{X}=\mathbf{x}} \perp^f \mathbf{Z}|\mathbb{U}$$

\square

A.10. Seção 5.1 (Identificabilidade na Descoberta Causal)

Lema A.24 (?). *Para quaisquer vértices $V_1, V_2 \in \mathcal{V}$ em um grafo causal, \mathcal{G} , as seguintes afirmações são equivalentes:*

1. V_1 e V_2 são adjacentes,
2. Não existe $\mathbb{V} \subseteq \mathcal{V} - \{V_1, V_2\}$ tal que $V_1 \perp V_2|\mathbb{V}$,
3. V_1 e V_2 não são d -separados dado $A := \text{Anc}(\{V_1, V_2\}) - \{V_1, V_2\}$,
4. V_1 e V_2 não são d -separados dado $Pa := Pa(\{V_1, V_2\}) - \{V_1, V_2\}$.

Demonstração. (1 \rightarrow 2) Sem perda de generalidade, suponha que $V_1 \rightarrow V_2$. Inicialmente, construímos uma f compatível com \mathcal{G} . Para todo $V \notin \{V_1, V_2\}$, tomamos $f(V|Pa(V)) = \mathbb{I}(V=0)$. Isto é, V é degenerado em 0. Além disso, tomamos $V_1|Pa(V_1) \sim \text{Bernoulli}(0.5)$ e $V_2|Pa(V_2) \equiv V_1$. Para todo $\mathbb{V} \subseteq \mathcal{V} - \{V_1, V_2\}$, $\mathbb{P}(\mathbb{V}=0) = 1$. Portanto, $\text{Cov}(V_1, V_2|\mathbb{V}) = \text{Cov}(V_1, V_2) = 0.25 \neq 0$. Portanto, V_1 e V_2 não são independentes dado \mathbb{V} segundo f . Como f é compatível com \mathcal{G} , decorre do [Teorema 2.49](#) que V_1 e V_2 não são d -separados dado \mathbb{V} .

(2 \rightarrow 3) Decorre do fato de que $\text{Anc}(\{V_1, V_2\}) - \{V_1, V_2\}$ é um caso particular de \mathbb{V} em (2).

(3 \rightarrow 4) Provaremos que se existe um caminho de V_1 em V_2 , C , que não está bloqueado dado A , ele também não está bloqueado dado Pa . Faremos esta prova em duas etapas: primeiramente considerando vértices em C que não sejam colisores e, a seguir, que sejam colisores. Tome um vértice em C , C_i , que não é um colisor. Como C não está bloqueado dado A , $C_i \notin A$. Como $Pa \subseteq A$, $C_i \notin Pa$. A seguir, tome um vértice em C , C_i , que é

um colisor. Como C não está bloqueado dado A , existe $V_1 \in A$ tal que V_1 é descendente de C . Como $V_1 \in A$, existe $V_2 \in Pa$ tal que $V_2 = V_1$ ou V_2 é descendente de V_1 . Portanto, $V_2 \in Pa$ é descendente de C_i . Decorre das conclusões anteriores que C não está bloqueado dado Pa .

(4 \rightarrow 1) Considere que V_1 e V_2 não são adjacentes e suponha por absurdo que há um caminho de V_1 a V_2 , C , não bloqueado dado Pa . Como V_1 e V_2 não são adjacentes, $|C| > 2$. Caso, $V_1 \leftarrow C_2$, então $C_2 \in Pa$ e C_2 não é um colisor, isto é, C está bloqueado dado Pa . Conclua que $V_1 \rightarrow C_2$. Por simetria, conclua que $C_{n-1} \rightarrow V_2$. Como $V_1 \rightarrow C_2$ e $C_{n-1} \rightarrow V_2$, há pelo menos um colisor em C . Defina C_i como o colisor de menor índice (o mais próximo de V_1) e C_j o de maior índice (o mais próximo de V_2). Por definição construção, C_i é descendente de V_1 e C_j é descendente de V_2 . Note que, se C_i é ascendente de V_2 e C_j é ascendente de V_1 , então

$$V_1 \rightarrow \dots \rightarrow C_i \rightarrow \dots \rightarrow V_2 \rightarrow \dots \rightarrow C_j \rightarrow \dots \rightarrow V_1,$$

é um ciclo em \mathcal{G} . Como \mathcal{G} é um DAG, ou C_i não é ascendente de V_2 ou C_j não é ascendente de V_1 . Sem perda de generalidade, considere que C_i não é ascendente de V_2 . Como C_i é descendente de V_1 , C_i também não é ascendente de V_1 . Como C_i não é ascendente de V_1 ou de V_2 , Não há vértice em Pa que é descendente de C_i . Portanto, C_i é um colisor e não há descendente de C_i em Pa . Conclua que C está bloqueado dado Pa , um absurdo. Portanto, $V_1 \perp V_2 | Pa$. \square

Lema A.25 (?). *Considere que \mathcal{G} é tal que V_1 é adjacente a V_2 , V_2 é adjacente a V_3 , mas V_1 não é adjacente a V_3 . Temos que $V_1 \rightarrow V_2$ e $V_2 \leftarrow V_3$ se e somente se V_1 não é d-separado de V_3 dado qualquer \mathbb{V} tal que $V_2 \in \mathbb{V}$.*

Demonstração. (\rightarrow) Considere que $V_1 \rightarrow V_2$ e $V_2 \leftarrow V_3$. Tome o caminho $C = V_1 \rightarrow V_2 \leftarrow V_3$. Como V_2 é um colisor em C , C não está bloqueado dado qualquer \mathbb{V} tal que $V_2 \in \mathbb{V}$. Conclua que V_1 não é d-separado de V_3 dado qualquer \mathbb{V} tal que $V_2 \in \mathbb{V}$.

(\leftarrow) Considere que $V_1 \leftarrow V_2$ ou $V_2 \rightarrow V_3$. Portanto, $V_2 \in Pa := Pa(\{V_1, V_3\}) - \{V_1, V_3\}$. Como V_1 não é adjacente a V_3 , conclua do [Lema A.24](#) que $V_1 \perp V_3 | Pa$. \square

Lema A.26. *Se \mathcal{G} e \mathcal{G}^* não tem o mesmo padrão, então \mathcal{G} e \mathcal{G}^* não são fielmente equivalentes.*

Demonstração. Decorre diretamente dos [Lemas A.24](#) e [A.25](#). \square

Lema A.27. *Se \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, $V_1 \rightarrow V_2 \leftarrow V_3$ é um colisor tanto em \mathcal{G} quanto em \mathcal{G}^* e Z é tal que $V_2 = Z$ ou Z é um descendente de V_2 em \mathcal{G} , então ou V_1 e V_3 são adjacentes em ambos os DAGs, ou existe um vértice V_* tal que $V_1 \rightarrow V_* \leftarrow V_3$ é um colisor em \mathcal{G}^* e $Z = V_*$ ou Z é descendente de V_* em \mathcal{G} .*

Demonstração. Para provar este resultado basta mostrar que, se V_1 e V_3 não são adjacentes nas condições do lema, então existe V_* com as condições especificadas. Como Z é descendente de V_2 em \mathcal{G} , existe um caminho direcionado em \mathcal{G} , $C = (C_1, \dots, C_n)$, tal que $C_1 = V_2$ e $C_n = Z$. Como \mathcal{G}^* tem o mesmo padrão de \mathcal{G} , C é um caminho de V_2 em Z . Faremos a demonstração por indução.

Suponha que $C_1 = V_2$ e $C_1 = Z$, isto é, $V_2 = Z$. Neste caso, $V_* = V_2$ satisfaz as condições do lema. A seguir, suponha que, se $n \leq n^*$, então existe V_* nas condições do lema e que $n = n^* + 1$. Se Z é descendente de V_2 em \mathcal{G}^* , então basta tomar $V_* = Z$. Caso contrário, existe um menor i tal que $C_i \leftarrow C_{i+1}$. Existem 2 casos a considerar: $i = 1$ e $i > 1$.

Se $i > 1$, então $C_{i-1} \rightarrow C_i \leftarrow C_{i+1}$ é um colisor em \mathcal{G}^* . Como eles formam uma cadeia em \mathcal{G} , e \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, C_{i-1} e C_{i+1} são adjacentes em ambos os DAGs. Como $C_{i-1} \xrightarrow{\mathcal{G}} C_i \xrightarrow{\mathcal{G}} C_{i+1}$ e \mathcal{G} é acíclico, $C_{i-1} \xrightarrow{\mathcal{G}} C_{i+1}$. Portanto, $C^* = (C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_n)$ é um caminho de tamanho $n - 1$ de V_2 a Z em \mathcal{G}^* que é um caminho direcionado em \mathcal{G} . Decorre da hipótese de indução que existe V_* tal qual desejado.

Se $i = 1$, $V_2 \xleftarrow{\mathcal{G}^*} C_2$. Portanto, $V_1 \rightarrow V_2 \leftarrow C_2$ e $V_3 \rightarrow V_2 \leftarrow C_2$ são colisores em \mathcal{G}^* , mas não em \mathcal{G} . Como \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, conclua que V_1 e C_2 , e V_3 e C_2 são adjacentes em ambos os DAGs. Como $V_1 \xrightarrow{\mathcal{G}} V_2 \xrightarrow{\mathcal{G}} C_2$ e \mathcal{G} é acíclico, $V_1 \xrightarrow{\mathcal{G}} C_2$. Por simetria $V_3 \xrightarrow{\mathcal{G}} C_2$. Portanto, $V_1 \rightarrow C_2 \leftarrow V_3$ é um colisor em \mathcal{G} . Como V_1 e V_3 não são adjacentes por suposição e \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, $V_1 \rightarrow C_2 \leftarrow V_3$ é um colisor em \mathcal{G}^* , (C_2, C_3, \dots, C_n) é um caminho de tamanho $n - 1$ em \mathcal{G}^* e que também é um caminho direcionado em \mathcal{G} . Portanto, pela hipótese de indução, existe V_* tal qual desejado. \square

Lema A.28. *Se \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, então \mathcal{G} e \mathcal{G}^* são fielmente equivalentes.*

Demonstração. Desejamos provar que para quaisquer $V_1, V_2 \in \mathcal{V}$ e $\mathbb{V} \subseteq \mathcal{V}$ tais que $V_1 \neq V_2$ e $\{V_1, V_2\} \cap \mathbb{V} = \emptyset$, se existe um caminho em \mathcal{G} de V_1 a V_2 não bloqueado dado \mathbb{V} , então existe um caminho em \mathcal{G}^* de V_1 a V_2 não bloqueado dado \mathbb{V} . Tome V_1, V_2 e \mathbb{V} arbitrários. Realizaremos a demonstração por indução.

Suponha que há um caminho de tamanho 2 de V_1 a V_2 em \mathcal{G} que não é bloqueado dado \mathbb{V} . Portanto, V_1 e V_2 são adjacentes em \mathcal{G} . Como \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, V_1 e V_2 são adjacentes em \mathcal{G}^* . Conclua do [Lema A.25](#) que (V_1, V_2) é um caminho não bloqueado dado \mathbb{V} em \mathcal{G}^* .

A seguir, suponha que, para todo caminho de tamanho menor ou igual a n em \mathcal{G} de V_1 a V_2 que não é bloqueado dado \mathbb{V} , existe um caminho em \mathcal{G}^* de V_1 a V_2 que não é bloqueado dado \mathbb{V} . Tome um caminho arbitrário de tamanho $n + 1$, $(C_1, C_2, \dots, C_n, C_{n+1})$ em \mathcal{G} de V_1 a V_2 que não é bloqueado dado \mathbb{V} . Desejamos provar que existe um caminho em \mathcal{G}^* de V_1 a V_2 que não é bloqueado dado \mathbb{V} .

Para tal, observamos inicialmente que, como \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, C é um caminho de V_1 a V_2 em \mathcal{G}^* . A seguir, dividiremos a análise em alguns casos:

- a) Considere que existe i tal que C_i é um colisor em \mathcal{G} e não é um colisor em \mathcal{G}^* . Como \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, C_{i-1} e C_{i+1} são adjacentes em \mathcal{G} . Sem perda de generalidade, suponha que $C_{i-1} \xrightarrow{\mathcal{G}} C_{i+1}$. Tome $C^* = (C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_{n+1})$. Com exceção de C_{i+1} , todos os vértices em C^* tem o mesmo tipo entre colisor e não-colisor que eles tem em C . Assim, todos os demais vértices estão abertos dado \mathbb{V} . Caso C_{i+1} não seja um colisor em C^* , então ele tem o mesmo tipo que em C e, assim, está aberto em C^* dado \mathbb{V} . A seguir, resta o caso em que C_{i+1} é um colisor em C^* . Como C_i é um colisor em C , C_i é descendente de C_{i+1} . Também, como C_i não estava bloqueado em C dado \mathbb{V} , existe $V \in \mathbb{V}$ tal que V é descendente de C_i . Conclua que V é descendente de C_{i+1} e, assim, C_{i+1} não está bloqueado em C^* dado \mathbb{V} . Portanto, C^* é um caminho de tamanho n de V_1 a V_2 em \mathcal{G} que não está bloqueado dado \mathbb{V} . Assim, esse caso está resolvido pela hipótese de indução.
- b) Considere que existe i tal que C_i não é um colisor em \mathcal{G} e é um colisor em \mathcal{G}^* . Como \mathcal{G} e \mathcal{G}^* tem o mesmo padrão, C_{i-1} e C_{i+1} são adjacentes em \mathcal{G} . Há 3 casos a analisar:
 - I) Caso $C_{i-1} \xrightarrow{\mathcal{G}} C_i \xrightarrow{\mathcal{G}} C_{i+1}$, como \mathcal{G} é um DAG, obtemos que $C_{i-1} \xrightarrow{\mathcal{G}} C_{i+1}$. Neste caso, todos os vértices no caminho $C^* = (C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_{n+1})$ tem o mesmo tipo entre colisor e não-colisor que eles tem em C . Assim, C^* é um caminho de tamanho n de V_1 a V_2 que não está bloqueado em \mathcal{G} dado \mathbb{V} . Assim, esse caso está resolvido pela hipótese de indução.
 - II) Caso $C_{i-1} \xleftarrow{\mathcal{G}} C_i \xleftarrow{\mathcal{G}} C_{i+1}$, a análise é análoga ao caso anterior.
 - III) Caso $C_{i-1} \xleftarrow{\mathcal{G}} C_i \xrightarrow{\mathcal{G}} C_{i+1}$, tome sem perda de generalidade que $C_{i-1} \xrightarrow{\mathcal{G}} C_{i+1}$. A seguir, há dois casos a considerar. Primeiramente, considere que C_{i-1} é um colisor em C em \mathcal{G} . Como C_i é um colisor em C em \mathcal{G}^* , C_{i-1} não é um colisor em C em \mathcal{G}^* . Portanto, C_{i-1} satisfaz as condições do caso tratado em a). A seguir, se C_{i-1} não é um colisor em C em \mathcal{G} , então todos os vértices em

$C^* = (C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_{n+1})$ tem o mesmo tipo entre colisor e não-colisor dado \mathbb{V} que em C . Portanto, C^* é um caminho de tamanho n de V_1 a V_2 que não está bloqueado dado \mathbb{V} em \mathcal{G} . Assim, esse caso está resolvido pela hipótese de indução.

- c) Resta o caso tal que, para todo i , tanto em \mathcal{G} quanto em \mathcal{G}^* , C_i tem o mesmo tipo em C entre colisor e não-colisor. Existem dois casos a analisar:
- I) Caso exista algum i tal que C_i é um colisor em C tanto em \mathcal{G} quanto em \mathcal{G}^* e C_{i-1} e C_{i+1} são adjacentes em \mathcal{G} . Sem perda de generalidade, suponha que $C_{i-1} \xrightarrow{\mathcal{G}} C_{i+1}$ e tome $C^* = (C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_{n+1})$. Todos os vértices em C^* com exceção de C_{i+1} tem o mesmo tipo entre colisor e não-colisor que em C . Caso C_{i+1} também tenha o mesmo tipo, então C^* não está bloqueado dado \mathbb{V} . Caso contrário, C_{i+1} é um colisor em C^* . Como C_i é um colisor em C , obtemos que $C_{i+1} \xrightarrow{\mathcal{G}} C_i$. Além disso, como C_i não está bloqueado em C dado \mathbb{V} , conclua que existe $V \in \mathbb{V}$ tal que V é descendente de C_i . Conclua das duas últimas sentenças que V é descendente de C_{i+1} . Assim, C_{i+1} não está bloqueado em C^* dado \mathbb{V} . Decorre que C^* é um caminho de tamanho n de V_1 a V_2 que não está bloqueado em \mathcal{G} dado \mathbb{V} . Assim, esse caso está resolvido pela hipótese de indução.
- II) Caso contrário, para todo i em que C_i é um colisor em C , C_{i-1} e C_{i+1} não são adjacentes. Portanto, decorre do [Lema A.27](#) que existe C_i^* tal que $C_{i-1} \xrightarrow{\mathcal{G}^*} C_i^* \xleftarrow{\mathcal{G}^*} C_{i+1}$ e existe $V \in \mathbb{V}$ que é descendente de C_i em \mathcal{G}^* . Para todo i em que C_i não é colisor em C , tome $C_i^* = C_i$. Por construção, C^* não está bloqueado em \mathcal{G}^* dado \mathbb{V} .

□

Prova do Teorema 5.5. Decorre dos [Lemas A.26](#) e [A.28](#)

□