

# Causality

Seminarium: Informatyka politycznie poprawna

*Fairness in Machine Learning - Limitations and Opportunities, Solon Barocas, Moritz Hardt, Arvind Narayanan*

Dominik Samorek, Klaudia Osowska

# Dlaczego causality jest przydatne?

Na samym początku spójrzmy na różnice między obserwacją a zdarzeniem

# Obserwacje a zdarzenia

Obserwacje oddają świat przez pryzmat pewnych cech.  
Dane, które mamy dzięki obserwacjom są "migawką" stanu świata.  
Dzięki obserwacjom możemy odpowiedzieć na pytanie:

Czy kierowcy, którzy mają 16 lat mają większy współczynnik wypadków niż kierowcy, którzy mają 18 lat?

# Obserwacje a zdarzenia

Obserwacje oddają świat przez pryzmat pewnych cech.  
Dane, które mamy dzięki obserwacjom są "migawką" stanu świata.  
Dzięki obserwacjom możemy odpowiedzieć na pytanie:

"Czy kierowcy, którzy mają 16 lat mają większy współczynnik wypadków niż kierowcy, którzy mają 18 lat?"

A co w przypadku pytań typu

"Czy ustanowienie mniejszej prędkości będzie dozwolone" dopiero od 18 roku życia sprawi, że w ruchu drogowym będzie mniej wypadków i będzie dozwolone?"

“Przytępowanie, że prawo jazdy będzie nie do wypadków spowodowanych 18-letnim” życia

Widząc efekt hipotetycznego pytania jak często występuje pewne zdarzenie, pytamy jaki

Nie jest tak łatwo znaleźć odpowiedź. Nawet jeśli starsi kierowcy mają niższy współczynnik wypadków, to może to być spowodowane większą liczbą lat doświadczenia.

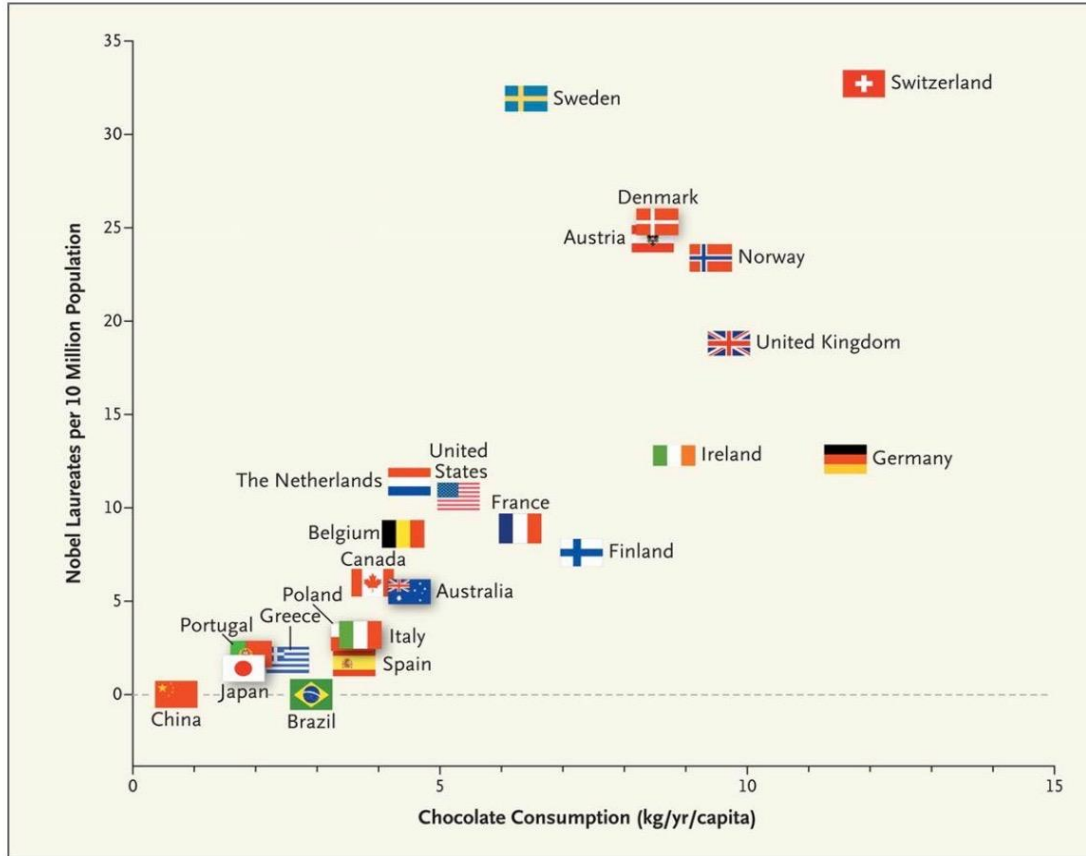
Dlaczego 18-letni kierowca z prawem jazdy od 2-miesiący miałby mieć mniejszą szansę na wypadek niż 16-letni kierowca z prawem jazdy od 2-miesiący?

“Przywstanie wiecie, że prawy jazdy lechnie dzwadowa dapienochod 18yok”życia

Jednak ci jesli grupy? na drodze jednej grupy wiekowej są zupełnie inne niż

Wzrostu i rozwoju w tym czasie kraje w dół i w górę pozostają i nie są w stanie porównać

# Przyczynowość a korelacja



F. H. Messerli: *Chocolate Consumption, Cognitive Function, and Nobel Laureates*, N Engl J Med 2012

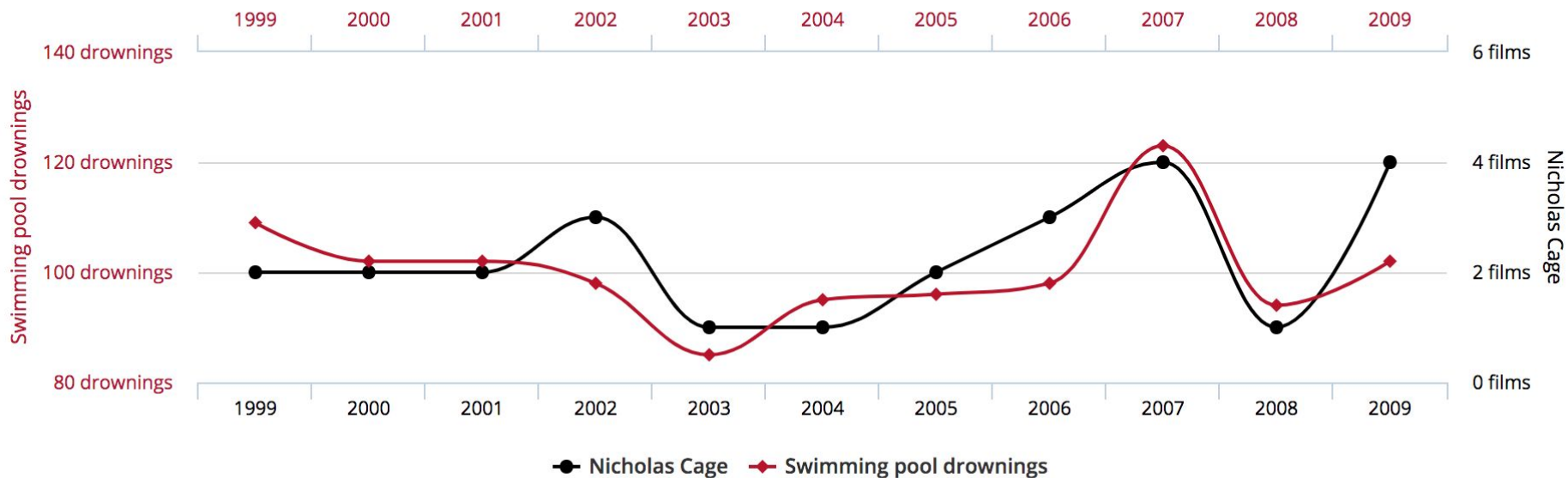
# Przyczynowość a korelacja



F. H. Messerli: *Chocolate Consumption, Cognitive Function, and Nobel Laureates*, N Engl J Med 2012

# Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% ( $r=0.666004$ )



tylervigen.com

Data sources: Centers for Disease Control & Prevention and Internet Movie Database

# Rozumowanie przyczynowe (ang. causal reasoning)

Framework, który pomaga zadawać pytania o efekt danego hipotetycznego zdarzenia.  
Wiedząc, że jest efekt zdarzenia możemy odwrócić pytanie i zapytać co spowodowało  
zdarzenie.  
Zatem ten framework pozwala w formalny sposób rozmawiać o przyczynach i

Jednak nie każde pytanie o przyczynę jest równie dobre:

Co spowodowało jego sukces?

zbyt ogólne



# Rozumowanie przyczynowe (ang. causal reasoning)

Rozumowanie przyczynowe daje nam język, który pozwala formalnie rozmawiać o przyczynach i skutkach, ale nie ułatwia wyboru odpowiednich pytań.

Nie ułatwia ono też zadania znalezienia odpowiedzi na pytania.

# Dane z przyjmowania kandydatów na Uniwersytet Kalifornijski, Berkeley 1973

- Było rozważanych 12763 osób do przyjęcia na jeden ze 101 wydziałów
- Z 4321 kobiet, które aplikowało zostało przyjętych 35%
- Z 8442 mężczyzn, którzy aplikowali zostało przyjętych 44%
- Biorąc pod uwagę tylko 6 największych wydziałów:

przyjętych zostało 44% mężczyzn i 30% kobiet

Table 7: UC Berkeley admissions data from 1973.

|            | Men     |              | Women   |              |
|------------|---------|--------------|---------|--------------|
| Department | Applied | Admitted (%) | Applied | Admitted (%) |
| A          | 825     | 62           | 108     | <b>82</b>    |
| B          | 520     | 60           | 25      | <b>68</b>    |
| C          | 325     | <b>37</b>    | 593     | 34           |
| D          | 417     | 33           | 375     | <b>35</b>    |
| E          | 191     | <b>28</b>    | 393     | 24           |
| F          | 373     | 6            | 341     | <b>7</b>     |

# Paradoks Simpsona

Paradoks Simpsona to zjawisko prawdopodobieństwa i statystyki, w którym trend pojawia się w kilku różnych grupach danych, ale znika lub odwraca się, gdy te grupy się łączą.

Gdy weźmiemy:

Y - zmienna mówiąca czy osoba przyjęta

A - zmienna mówiąca czy osoba jest płci żeńskiej

Z - zmienna losowa mówiąca na jaki wydział osoba przyjęta

To można zapisać:

1.  $\mathbb{P}\{Y \mid A\} < \mathbb{P}\{Y \mid \neg A\}$
2.  $\mathbb{P}\{Y \mid A, Z = z\} > \mathbb{P}\{Y \mid \neg A, Z = z\}$  for all values  $z$  that the random variable  $Z$  assumes.

# Paradoks Simpsona

|               | Operacje      | Małoinwazyjne procedury |
|---------------|---------------|-------------------------|
| Małe kamienie | 93% (81/87)   | 87% (234/270)           |
| Duże kamienie | 73% (192/263) | 69% (55/80)             |
| Razem         | 78% (273/350) | 83% (289/350)           |

źródło: <https://www.crazynauka.pl/rozmiar-ma-znaczenie-czyli-paradoks-simpsona/>

- Płeć ma wpływ na wybór wydziału
- Kobiety aplikują na wydziały, na których jest większa konkurencja, zatem zostają odrzucane częściej niż mężczyźni

Table 7: UC Berkeley admissions data from 1973.

| Department | Men     |              | Women   |              |
|------------|---------|--------------|---------|--------------|
|            | Applied | Admitted (%) | Applied | Admitted (%) |
| A          | 825     | 62           | 108     | <b>82</b>    |
| B          | 520     | 60           | 25      | <b>68</b>    |
| C          | 325     | 37           | 593     | 34           |
| D          | 417     | 33           | 375     | 35           |
| E          | 191     | <b>28</b>    | 393     | 24           |
| F          | 373     | 6            | 341     | 7            |

*The bias in the aggregated data stems not from any pattern of discrimination on the part of admissions committees, which seems quite fair on the whole, but apparently from prior screening at earlier levels of the educational system. Women are shunted by their socialization and education toward fields of graduate study that are generally more crowded, less productive of completed degrees, and less well funded, and that frequently offer poorer professional employment prospects.<sup>125</sup>*

<sup>125</sup> Bickel, Hammel, O'Connell, and others, "Sex Bias in Graduate Admissions."

- Dlaczego kobiety aplikowały na bardziej konkurencyjne wydziały?

# Causal models

Można o nich myśleć jako o programach, które poprzez ciąg instrukcji generują z zakłócających zmiennych pewien rozkład.

Jednak tutaj każda instrukcja jest w języku matematyki zamiast w języku programowania.

# Przykład

Weźmy hipotetyczną populację w której:

- 1) Każdy regularnie ćwiczy z prawdopodobieństwem  $\frac{1}{2}$
- 2) Każdy z prawdopodobieństwem  $\frac{1}{3}$  ma ukrytą skłonność do nadwagi, która ukazuje się w konsekwencji braku regularnych ćwiczeń
- 3) Każdy z prawdopodobieństwem  $\frac{1}{3}$  ma ukrytą wadę serca, która również objawia się przy braku regularnych ćwiczeń

Niech

X - zmienna mówiąca czy ktoś regularnie ćwiczy

W - zmienna mówiąca czy ktoś ma nadwagę

H - zmienna mówiąca czy ktoś ma wadę serca

X - zmienna mówiąca czy ktoś regularnie ćwiczy

W - zmienna mówiąca czy ktoś ma nadwagę

H - zmienna mówiąca czy ktoś ma wadę serca

# Przykład - causal model

1. Weźmy trzy zmienne zakłócające o rozkładzie Bernoulliego

$$U1 \sim B(1/2), U2 \sim B(1/3), U3 \sim B(1/3)$$

2.  $X := U1$
3.  $W := \text{if } X = 1 \text{ then } 0 \text{ else } U2$
4.  $H := \text{if } X = 1 \text{ then } 0 \text{ else } U3$

X - zmienna mówiąca czy ktoś regularnie ćwiczy

W - zmienna mówiąca czy ktoś ma nadwagę

H - zmienna mówiąca czy ktoś ma wadę serca

# Przykład - causal model

1. Weźmy trzy zmienne losowe o rozkładzie Bernoulliego

$$U1 \sim B(1/2), U2 \sim B(1/3), U3 \sim B(1/3)$$

2.  $X := U1$

3.  $W := \text{if } X = 1 \text{ then } 0 \text{ else } U2$

4.  $H := \text{if } X = 1 \text{ then } 0 \text{ else } U3$

- Ćwiczenia zapobiegają wadom serca oraz nadwadze. W wypadku braku ćwiczeń wady serca i nadwaga występują od siebie niezależnie.
- Model generuje łączny rozkład (ang. joint distribution) dla zmiennych X, W, H.
- Można obliczyć prawdopodobieństwa przy tym rozkładzie:

$$P(H=1) = 1/2 * 1/3 = 1/6, P(H=1 | W=1) = 1/3$$

Czy to oznacza, że nadwaga powoduje choroby serca?

# Czy nadwaga powoduje choroby serca?

Weźmy nowy program:

2.  $X := U1$

3.  $W := 1$

4.  $H := \text{if } X = 1 \text{ then } 0 \text{ else } U3$

Ten model generuje już inny rozkład

$P(H=1) = \frac{1}{6}$

Czyli sprawienie, że osoba na pewno będzie miała nadwagę, nie wpływa na prawdopodobieństwo wystąpienia wady serca!

Wniosek:  **$W = 1$  to nie to samo co  $W := 1$**

$W = 1$  to obserwacja, a  $W := 1$  to interwencja, akcja

# Rozważmy inną populację

2.  $W := U2$

3.  $X := \text{if } W = 0 \text{ then } 0 \text{ else } U1$

4.  $H := \text{if } X = 1 \text{ then } 0 \text{ else } U3$

Przy  $W := 1$  prawdopodobieństwo choroby serca wynosi  $\frac{1}{6}$

Przy  $W = 1$  również prawdopodobieństwo choroby serca wynosi  $\frac{1}{6}$

Zatem podstawienie “ $:=$ ” może odpowiadać prawdopodobieństwu warunkowemu, ale nie musi.

“ $:=$ ” odpowiada akcji, czasem nazywamy go do-operatorem i zapisujemy **do( $W := 1$ )**

# Causal model formalnie

**Causal model  $M$**  dany jest przez zbiór zmiennych  $X_1, \dots, X_d$  oraz zbiór przypisań w postaci

$$X_i := f_i(P_i, U_i), \quad i = 1, \dots, d$$

gdzie  $P_i \subseteq \{X_1, \dots, X_d\}$  i zmienne ze zbioru  $P_i$  nazywamy rodzicami  $X_i$ ,

a zmienne losowe  $U_1, \dots, U_d$  nazywamy zmiennymi zakłócającymi i muszą one być wzajemnie niezależne

Graf skierowany odpowiadający modelowi posiada dokładnie jeden węzeł dla każdej zmiennej  $X_i$  i każdy z tych węzłów ma krawędzie wchodzące od rodziców danego węzła. Taki graf nazywamy causal graph.

O zmiennych zakłócających można myśleć jako o czynnikach zewnętrznych.

Rodzice pewnego wężła odpowiadającego zmiennej  $X_i$  są nazywane *bezpośrednią przyczyną*  $X_i$ . Podobnie  $X_i$  jest nazywane *bezpośrednim skutkiem* dla każdego z rodziców tego wężła.

Causal model to zbiór pewnych *założeń* co do tego jak konkretne zmienne na siebie nawzajem oddziałują i każde przypisanie związane jest z response function.

**Causal model M** dany jest przez zbiór zmiennych  $X_1, \dots, X_d$  oraz zbiór przypisań w postaci

$$X_i := f_i(P_i, U_i), \quad i = 1, \dots, d$$

gdzie  $P_i \subseteq \{X_1, \dots, X_d\}$  i zmienne ze zbioru  $P_i$  nazywamy rodzicami  $X_i$ ,

a zmienne losowe  $U_1, \dots, U_d$  nazywamy zmiennymi zakłócającymi i muszą one być wzajemnie niezależne

Graf skierowany odpowiadający modelowi posiada dokładnie jeden węzeł dla każdej zmiennej  $X_i$  i każdy z tych węzłów ma krawędzie wchodzące od rodziców danego wężła. Taki graf nazywamy causal graph.

To jak faktycznie nasze założenia mają się do rzeczywistości to inna kwestia.

Dlatego kiedy wyciągamy wnioski o skutkach i przyczynach w oparciu o model, nie musi to oznaczać, że takie związki faktycznie zachodzą w rzeczywistości.

Wprowadzamy też oznaczenie na prawdopodobieństwo przy danym rozkładzie:

M - model, E - zdarzenie

$$P_M\{E\}$$

Przykładowo przy wcześniej rozważanej populacji danej przez model M:

2.  $X := U1$

3.  $W := \text{if } X = 1 \text{ then } 0 \text{ else } U2$

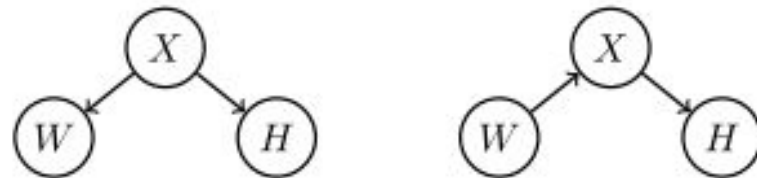
4.  $H := \text{if } X = 1 \text{ then } 0 \text{ else } U3$

Możemy zapisać  $P_M\{H\} = \frac{1}{6}$

# Causal graphs

Strukturalne modele przyczynowe (*structural causal models*) możemy reprezentować bardziej ogólnie przy pomocy **grafów**.

Nie interesują nas tutaj dokładne wzory na każdą zmienną, modelujemy **strukturę zależności**.



Grafy obrazujące wcześniejsze przykłady modeli chorób serca

# Motywacja

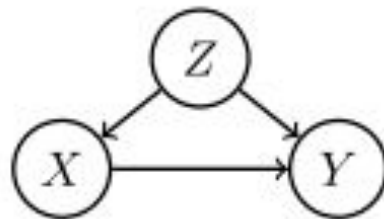
- Chcemy umieć odpowiadać na hipotetyczne pytania:
  - Czy odsetek ludzi z chorobą serca wzrósłby gdyby wszyscy mieli nadwagę?
- Dysponujemy danymi w naszej rzeczywistości
  - rozkład warunkowy to nie to czego szukamy, czai się w nim wiele nieprawdziwych zależności
- Operator *do* mówi o rozkładzie danych z tej hipotetycznej rzeczywistości, nie mamy takich danych!
- Żeby uzyskać prawdziwy *przyczynowy* wpływ danej zmiennej, musimy pozbyć się tych nieprawdziwych zależności

# Dalszy plan prezentacji

- Nieprawdziwe zależności w grafach
- Co interwencja robi z grafem
- Jak próbkować z rozkładu interwencyjnego
- Jak wyliczyć rozkład interwencyjny z oryginalnych danych
- Pozbywanie się nieprawdziwych zależności

# Forks

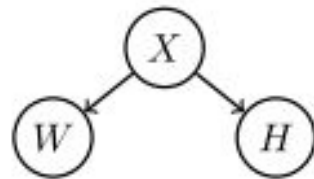
- Z jest wspólną przyczyną  $X$  i  $Y$
- Wykonywanie ćwiczeń było przykładem forka w modelu chorób serca
- **Efekt mylenia** (*confounding effect*)



1. Weźmy trzy zmienne zakłócające o rozkładzie Bernoulliego

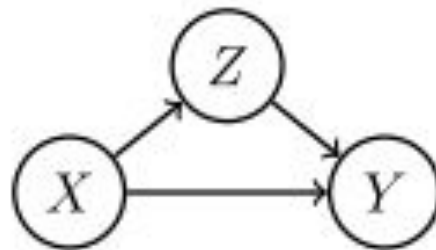
$$U1 \sim B(\frac{1}{2}), U2 \sim B(\frac{1}{3}), U3 \sim B(\frac{1}{3})$$

2.  $X := U1$
3.  $W := \text{if } X = 1 \text{ then } 0 \text{ else } U2$
4.  $H := \text{if } X = 1 \text{ then } 0 \text{ else } U3$



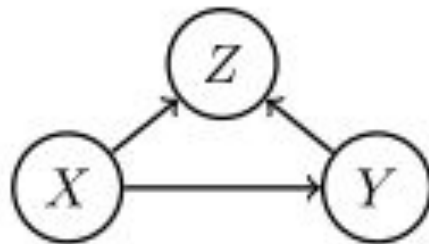
# Mediators

- Z ma udział we wpływie  $X$  na  $Y$
- Przykład z rekrutacją na studia: wybór wydziału pośredniczył we wpływie płci na wynik rekrutacji
- Bardzo ważny czynnik w analizie dyskryminacji (widać w przykładzie), ale szczegóły w dalszej części książki



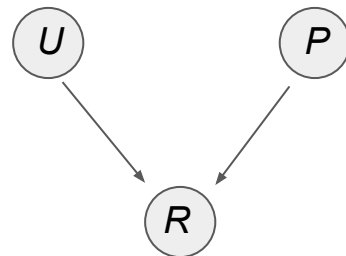
# Colliders

- $X$  i  $Y$  mogą nie być zmylone
- Musimy uważać, gdy warunkujemy *collider*
- *Explaining away effect, collider bias, Berkson's law*
- Negatywna korelacja między  $X$  a  $Y$



# Collider bias - Ellenberg example

- Alicja chodzi na randki  $R=1$  tylko z mężczyznami, których uprzejmość  $U$  plus przystojność  $P$  przekracza pewien próg
- **$R$  to collider**
- W takim razie uprzejmy mężczyzna, z którym Alicja mogłaby randkować nie potrzebuje być aż tak bardzo przystojny. I odwrotnie
- Zmienne  $U \mid R=1$  i  $P \mid R=1$  są zależne i ujemnie skorelowane!



# Akcja/Interwencja i *do-operator*

W strukturalnym modelu przyczynowym mamy podstawienia postaci:

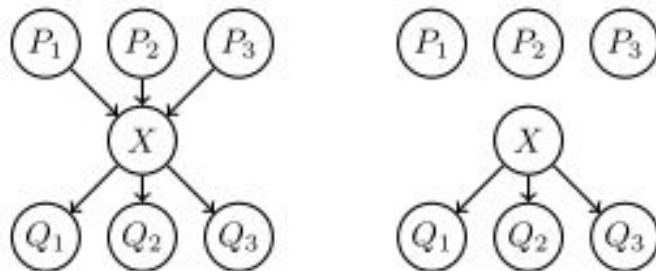
$$X := f(P, U)$$

Naszą operację interwencji oznaczamy przez:

$$X := x$$

Interesuje nas:

$$\mathbf{P}\{E \mid \text{do}(X := x)\} = \mathbf{P}_{M[X:=x]}(E)$$



# Efekt przyczynowy

- Efekt przyczynowy akcji  $X:=x$  na zmienną  $Y$  odnosi się do rozkładu  $Y$  w modelu  $M[X:=x]$
- Efekt przyczynowy zmiennej  $X$  na zmienną  $Y$  odnosi się do tego jak wszystkie możliwe podstawienia  $X$  mogą wpłynąć na rozkład  $Y$
- W przypadku binarnego  $X$  może nas interesować wartość:

$$\mathbf{E}_{M[X:=1]}[Y] - \mathbf{E}_{M[X:=0]}[Y] .$$

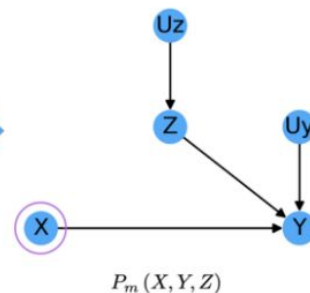
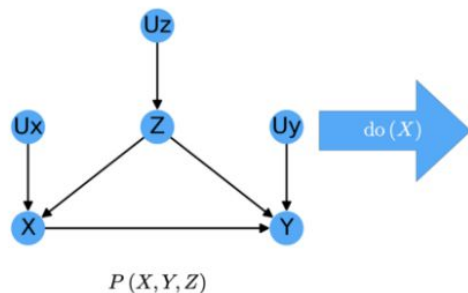
*(average treatment effect)*

# Rozkład interwencyjny

- Randomizowane kontrolowane badania kliniczne (ang. *randomized controlled trials*)
  - dzielimy badanych losowo na dwie grupy
  - ani pacjenci ani lekarze nie wiedzą kto jest w której grupie
  - jednej grupie podajemy lek, drugiej podajemy placebo
  - Takie badania pozwalają nam uzyskać próbki bezpośrednio z rozkładu zmodyfikowanego przez interwencję!
- 
- Niestety nie zawsze jest możliwość wykonania takich badań, wolelibyśmy umieć policzyć wpływ akcji z “normalnych” danych

# Rozkład interwencyjny - leczenie kamieni nerkowych

- $X$  - rodzaj leczenia
- $Z$  - wielkość kamienia
- $Y$  - sukces leczenia



$$P\{Y \mid do(X := x')\} = P_m\{Y\}$$

$$P_m\{X = x'\} = 1$$

$$P_m\{Z\} = P\{Z\}$$

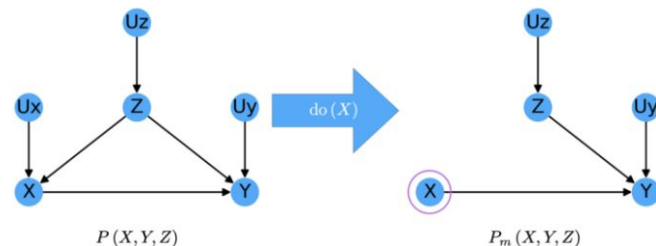
$$P_m\{Y \mid X, Z\} = P\{Y \mid X, Z\}$$

$$P_m\{Y\} = \sum_x \sum_z P_m\{Y \mid X = x, Z = z\} P_m\{Z = z \mid X = x\} P_m\{X = x\}$$

$$P\{Y = y \mid do(X := x')\} = \sum_z P\{Y = y \mid X = x', Z = z\} P\{Z = z\}$$

# Leczenie kamieni nerkowych

- $X$  - rodzaj leczenia
- $Z$  - wielkość kamienia
- $Y$  - sukces leczenia



$$E_{M[X:=operacja]}[Y] =$$

$$P\{Y = 1 \mid do(X := operacja)\} =$$

$$\sum_z P\{Y = 1 \mid X = operacja, Z = z\}P\{Z = z\} = 0.832 > 0.782$$

|               | Operacje      | Małoinwazyjne procedury |
|---------------|---------------|-------------------------|
| Małe kamienie | 93% (81/87)   | 87% (234/270)           |
| Duże kamienie | 73% (192/263) | 69% (55/80)             |
| Razem         | 78% (273/350) | 83% (289/350)           |

# *The Adjustment Formula* a prawdopodobieństwo warunkowe

- Możemy obliczyć wpływ akcji  $do(X:=x)$  kontrolując wszystkich rodziców  $X$  (zbiór  $Z$ ):

$$P\{Y = y \mid do(X := x)\} = \sum_z P\{Y = y \mid X = x, Z = z\}P\{Z = z\}$$

- Dla przypomnienia wzór na prawdopodobieństwo warunkowe wygląda tak:

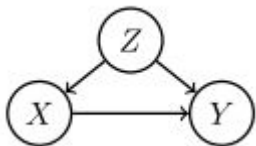
$$P\{Y = y \mid X = x\} = \sum_z P\{Y = y \mid X = x, Z = z\}P\{Z = z \mid X = x\}$$

# *Backdoor Criterion*

- Możemy obliczyć prawdopodobieństwo zdarzeń pod warunkiem akcji  $do(X:=x)$  kontrolując inny zbiór niż rodzice  $X$
- Zbiór  $Z$  spełnia *Backdoor criterion* jeśli
  - żaden z elementów  $Z$  nie jest potomkiem  $X$
  - element  $Z$  **blokuje** każdą nieskierowaną ścieżkę między  $X$  a  $Y$ , która zaczyna się krawędzią wchodzącą do  $X$
- W skrócie: chcemy zostawić wszystkie ścieżki *przyczynowe* z  $X$  do  $Y$ , ale zablokować wszystkie inne

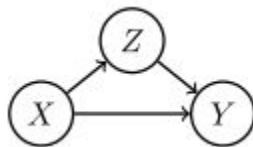
# Przyczynowa vs. nieprzyczynowa zależność

Fork



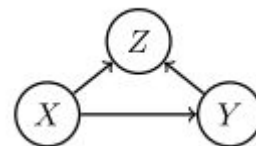
Kontrolowanie Z usuwa  
efekt zmylenia  
(*deconfounding*)

Mediator



Kontrolowanie Z  
usuwa część  
przyczynowego  
wpływu X na Y

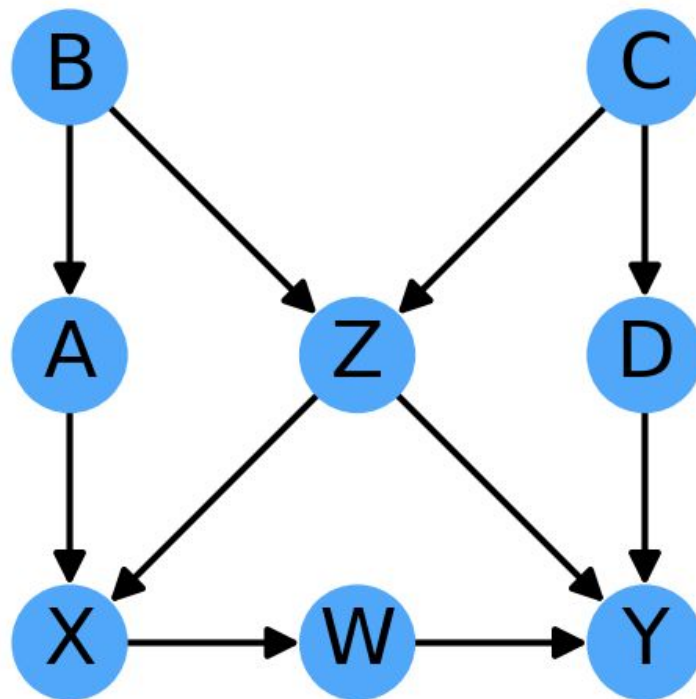
Collider



Kontrolowanie Z  
wprowadza  
korelację między  
X a Y. To znaczy  
tworzy nowy  
kanał do  
*przepływu*  
*informacji!*

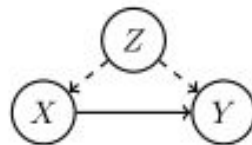
# Jak zablokować ścieżkę?

- Chcemy zbadać wpływ zmiennej  $X$  na  $Y$
- Musimy kontrolować  $Z$  (collider)
- To otwiera nam przepływ informacji
- Żeby go zablokować musimy kontrolować jeszcze jedną ze zmiennych  $A, B, C$  lub  $D$

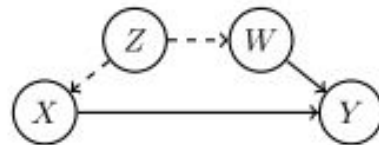


# Nie w pełni zaobserwowany model

- Czasem może nie udać nam się zaobserwować wszystkich zmiennych w modelu
- Nie zawsze jest to problem



Przykład z  
rakiem płuc



Chcąc ustalić wpływ  $X$   
na  $Y$  bez  
zaobserwowanego  $Z$   
wystarczy kontrolować  
 $W$

# Podsumowanie

- Rozumowanie przyczynowe daje nam matematyczne narzędzia do opisywania przyczyn i skutków zdarzeń
- Akcja/interwencja jest czymś zupełnie innym od obserwacji!
- W wielu przypadkach możemy ustalić skutki zdarzeń przy użyciu danych, którymi dysponujemy (przykład z chorobami serca)
- Czasem potrzebujemy jednak specjalnych eksperymentów lub badań (sprawdzanie skuteczności leku)
- Zwykle największym problemem jest zawarcie wszystkich potrzebnych zmiennych w naszym modelu
- Błędne wnioski możemy wyciągnąć przez brak “zewnętrznych” zmiennych (wspólna przyczyna jedzenia czekolady i zostania noblistą) lub “wewnętrznych” (rekrutacja na studia w Berkely)

# Materiały warte polecenia

- [Lectures on Causality: Jonas Peters MIT](#)
- [Cykl artykułów "Causal Inference In Statistics" z medium.com](#)
- [Causality, Judea Pearl](#)