

# Budowa modeli decyzyjnych z danych

**Marek J. Drużdżel**

**Politechnika Białostocka**

**Wydział Informatyki**

***m.druzdzel@pb.edu.pl***

***http://www.wi.pb.edu.pl/~druzdzel/***

# Plan wykładu

- **Wstęp: statystyka bayesowska**
- **Przyczynowość i prawdopodobieństwo**
- **Uczenie przyczynowości z danych**
- **Przykład**
- **Demo**
- **Uwagi końcowe**



# Bayesowska teoria prawdopodobieństwa






# Twierdzenie Bayesa i statystyka Bayesa

- Wstęp: statystyka bayesowska
- Przyczynowość i prawdopodobieństwo
- Uczenie przyczynowości z danych
- Przykład
- Demo
- Uwagi końcowe

Wszechstronne podejście, które wydaje się rozwiązywać różnorodne problemy, pochodzące od osiemnastowiecznego angielskiego matematyka, wielebnego Thomasa Bayesa

([http://en.wikipedia.org/wiki/Thomas\\_Bayes](http://en.wikipedia.org/wiki/Thomas_Bayes))



the theory   
that would   
not die   
how bayes' rule cracked   
the enigma code,  
hunted down russian  
submarines & emerged  
triumphant from two   
centuries of controversy  
sharon bertsch mcgrayne

Teoria Bayesa jest tak „na czasie”, że popularno-naukowa książka „Teoria, która nie chciała umrzeć”, wydana w 2011 roku, stała się bestsellerem

**Modelowanie bayesowskie jest niezawodne i rozwiązuje trudne problemy.**

**Może wykorzystywać zarówno dane, jak i wiedzę ekspercką.**

Polecany wykład:

<http://www.youtube.com/watch?v=8oD6eBkjF9o>

Budowa modeli decyzyjnych z danych

# Twierdzenie Bayesa

Łatwe do udowodnienia twierdzenie, otrzymane w następujący sposób:

Z definicji prawdopodobieństwa warunkowego;

$$P(A|B) = P(A,B) / P(B) \text{ i } P(B|A) = P(A,B) / P(A)$$

mamy

$$P(A|B) P(B) = P(A,B) \text{ i } P(B|A) P(A) = P(A,B)$$

a więc

$$P(A|B) P(B) = P(B|A) P(A)$$

czyli

$$P(A|B) = P(B|A) / P(B) P(A)$$

Prawdopodobieństwo a-posteriori

Prawdopodobieństwo a-priori

Twierdzenie Bayesa daje nam mechanizm zmiany opinii w świetle nowych obserwacji!

## Jaki jest związek statystyki bayesowskiej ze statystyką klasyczną?



**Klasyccy statystycy:** „Nie mamy pojęcia 😊. Prawdopodobieństwo jest ograniczającą częstotliwością. Wojna nuklearna nie jest powtarzalnym procesem ”.

**Bayesowcy:** „0.24 😊. Prawdopodobieństwo jest miarą przekonania”

# Przykład użycia twierdzenia Bayesa

Niech częstość występowania kiły w populacji młodych ludzi planujących zawarcie małżeństwa w Pensylwanii wyniesie 0,001.

Niech (obowiązkowy) test wymagany do uzyskania aktu małżeństwa będzie miał czułość 0,98 i swoistość 0,95.

Jakie jest prawdopodobieństwo, że twojanarzeczona, u której wykryto kiłę, ma kiłę?

$$P(S|+) = P(+|S)/P(+)$$

(twierdzenie Bayesa)

$$P(+)= P(+|S) P(S) + P(+|\sim S) P(\sim S)$$

(twierdzenie prawdopodobieństwa całkowitego)

$$P(+)= 0.98 \cdot 0.001 + 0.05 \cdot 0.999 = 0.05093$$

$$P(S|+) = 0.98 \cdot 0.001 / 0.05093 \cdot 0.001$$

Prawdopodobieństwo a-posteriori

Prawdopodobieństwo a-priori

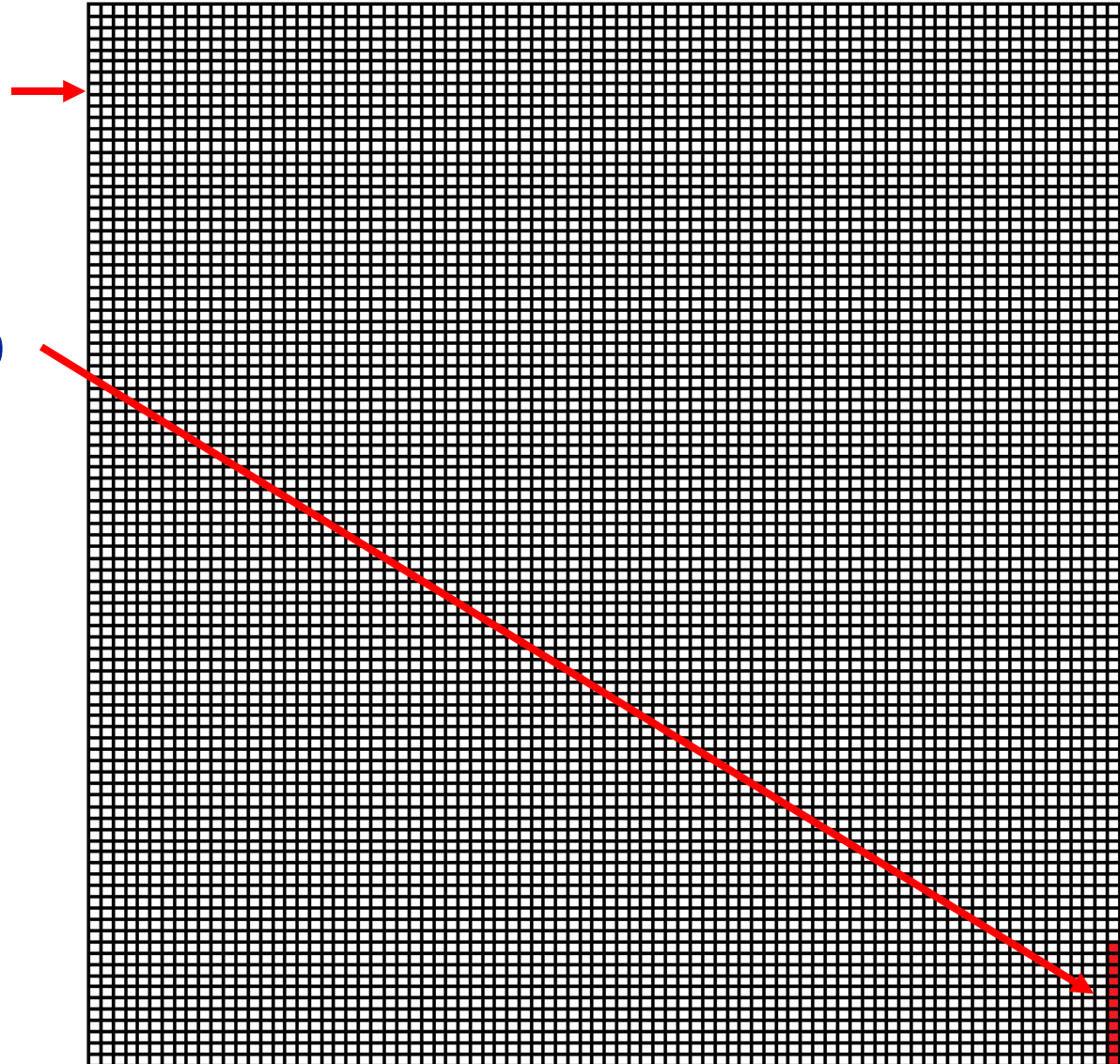
**0.01924**

# Lepszy interfejs człowieka dla tego samego problemu

Wyobraźmy sobie populację 10 000 osobników.

Częstość wynosząca 0,001 oznacza, że chorych jest 10 na 10 tysięcy.

Poddamy ich wszystkich testowi.





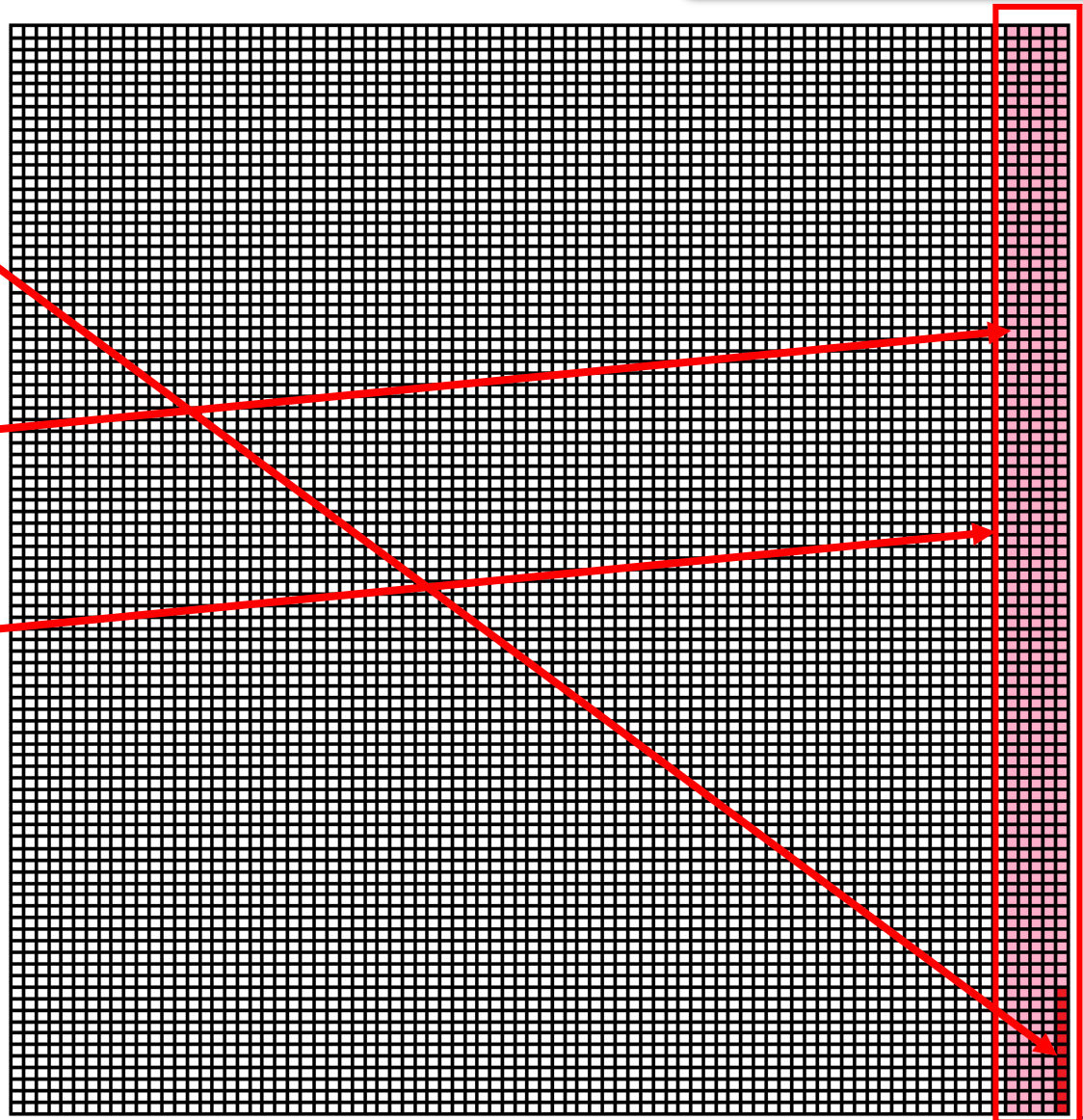
## Lepszy interfejs człowieka dla tego samego problemu

Przy czułości 98%, 9,8 z 10 chorych zostanie prawidłowo wykrytych.

Przy swoistości 95% otrzymamy 5% (z 9990), co daje 499,5 fałszywych trafień.

Wśród wszystkich osób, które uzyskały wynik pozytywny, około  $9,8 / (9,8 + 499,5) \approx 2\%$  będzie chorych.

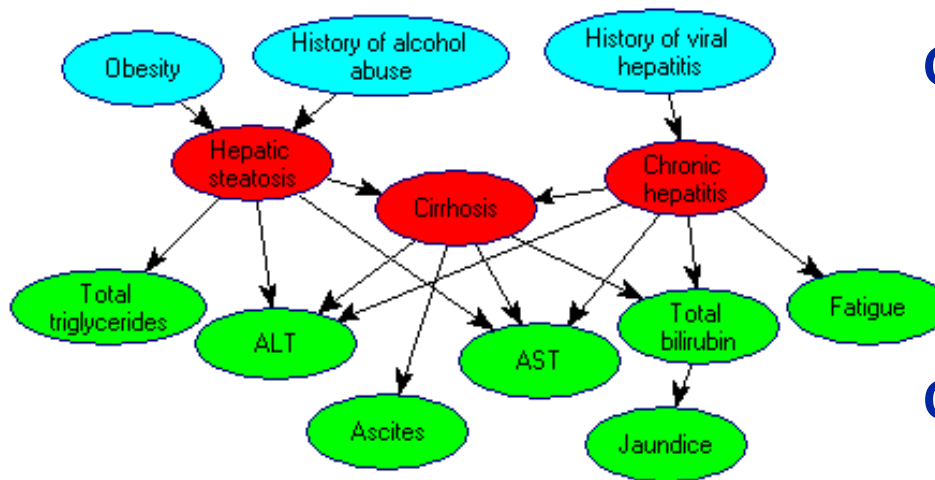
**Czy to jest łatwiej zrozumieć 😊?**



# Sieci bayesowskie

# Sieci bayesowskie

Sieć bayesowska (nazywana również siecią przekonań, siecią probabilistyczną lub siecią przyczynową) jest acyklicznym grafem skierowanym (DAG) składającym się z:



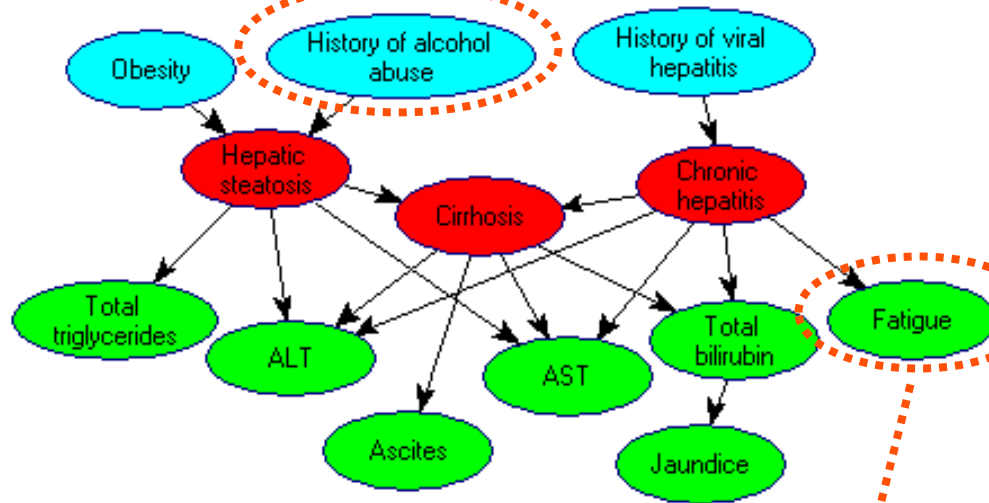
Część **jakościowa**, kodująca zmienne domeny (węzły) oraz probabilistyczne (zwykle przyczynowe) wpływy między nimi (łuki).

Część **ilościowa**, kodująca łączny rozkład prawdopodobieństwa dla tych zmiennych.

# Sieci bayesowskie: Parametry numeryczne

|         |      |
|---------|------|
|         |      |
| present | 0.15 |
| absent  | 0.85 |

Tabele rozkładu prawdopodobieństwa **a-priori** dla węzłów bez poprzedników (*History of viral hepatitis, History of alcohol abuse, Obesity*)

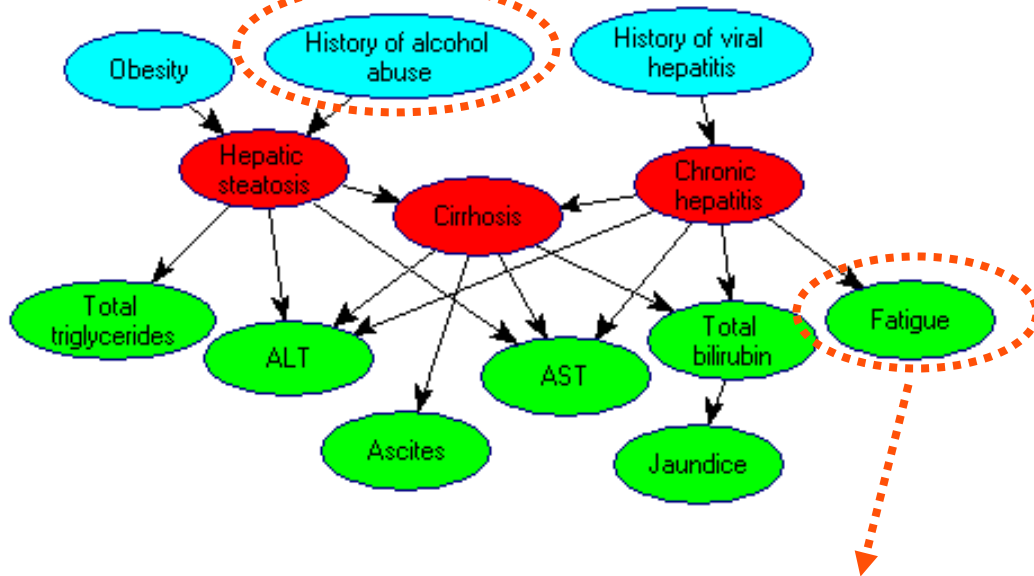


Tabele rozkładów prawdopodobieństwa **warunkowego** dla węzłów z poprzednikami (*Fatigue, Jaundice, ...*)

| Chronic hepatitis | present | absent |
|-------------------|---------|--------|
| present           | 0.6     | 0.2    |
| absent            | 0.4     | 0.8    |

# Skąd pochodzą parametry sieci bayesowskich?

|         |      |
|---------|------|
|         |      |
| present | 0.15 |
| absent  | 0.85 |



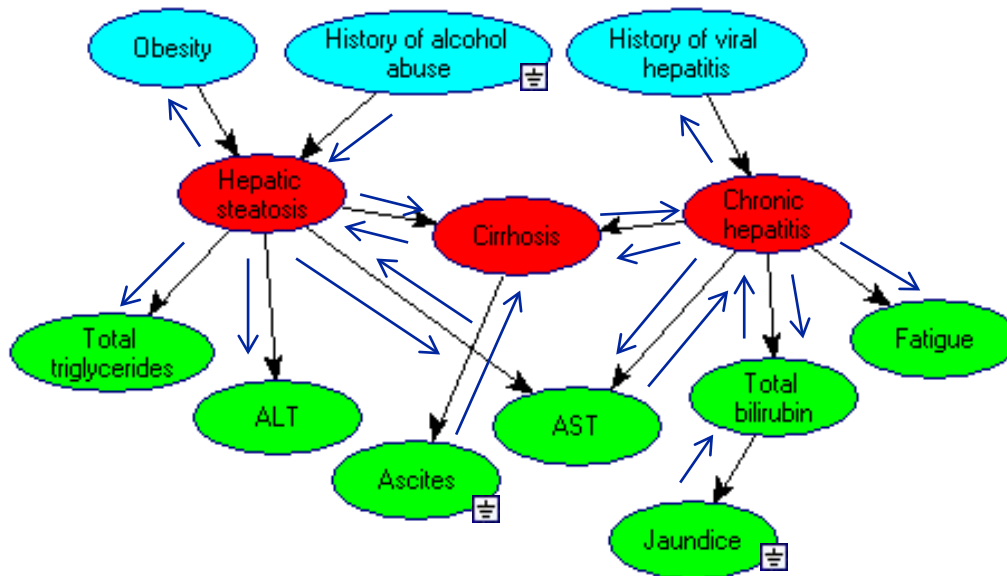
- Podręczniki
- Literatura
- Opinia eksperta
- Bazy danych

| Chronic hepatitis | present | absent |
|-------------------|---------|--------|
| present           | 0.6     | 0.2    |
| absent            | 0.4     | 0.8    |

# Obliczenia w sieciach bayesowskich

- Wstęp: statystyka bayesowska
- Przyczynowość i prawdopodobieństwo
- Uczenie przyczynowości z danych
- Przykład
- Demo
- Uwagi końcowe

Najważniejszym typem obliczeń w sieciach bayesowskich jest aktualizacja prawdopodobieństwa hipotezy (np. diagnozy) na podstawie nowych obserwacji (np. symptomy, wyniki testów).



**Przykład:**

Jakie jest prawdopodobieństwo przewlekłego zapalenia wątroby u alkoholika z żółtaczką i pajączakami?

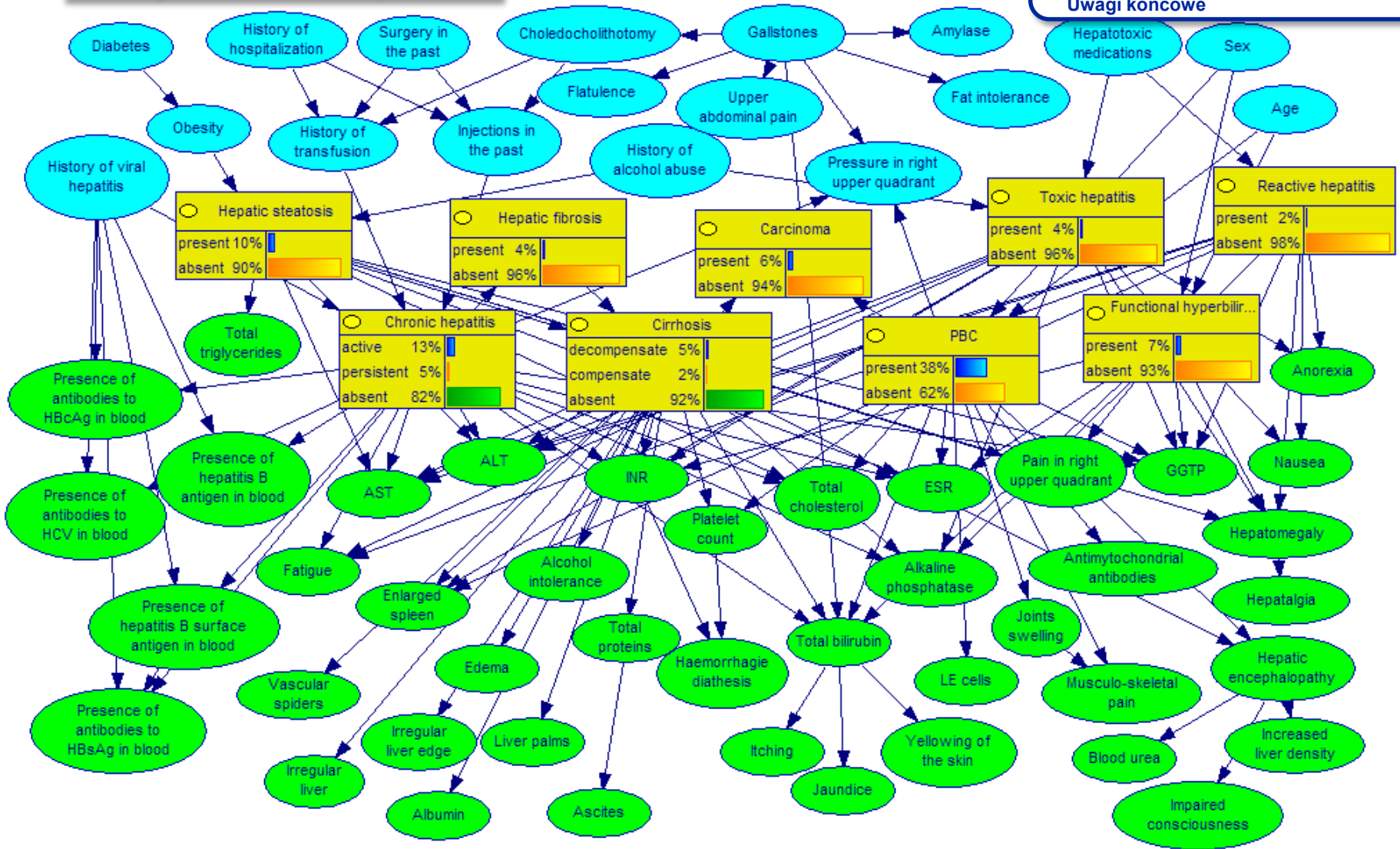
Która choroba jest najbardziej prawdopodobna?

Jakie testy powinniśmy wykonać w następnej kolejności?

$P(\text{Hepatitis} \mid \text{alcoholism}=\text{present}, \text{jaundice}=\text{present}, \text{ascites}=\text{present})?$

# Przykład: Hepar II

- Wstęp: statystyka bayesowska
- Przyczynowość i prawdopodobieństwo
- Uczenie przyczynowości z danych
- Przykład
- Demo
- Uwagi końcowe

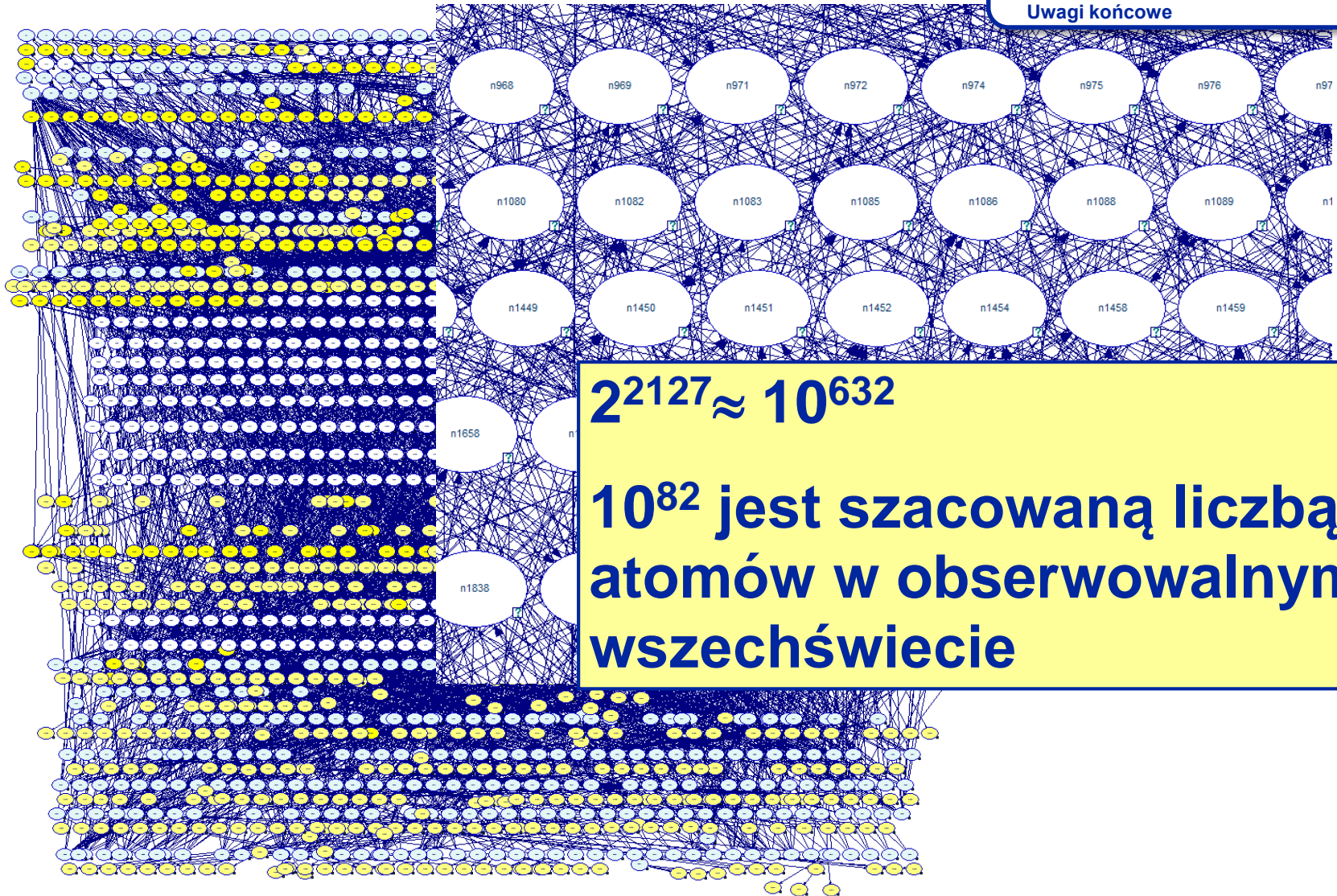


70 zmiennych; 2.139 parametrów numerycznych (zamiast ponad  $2^{70} \approx 10^{21}$ !)



# Diagnoza lokomotyw Diesla

- Wstęp: statystyka bayesowska
- Przyczynowość i prawdopodobieństwo
- Uczenie przyczynowości z danych
- Przykład
- Demo
- Uwagi końcowe



$$2^{2127} \approx 10^{632}$$

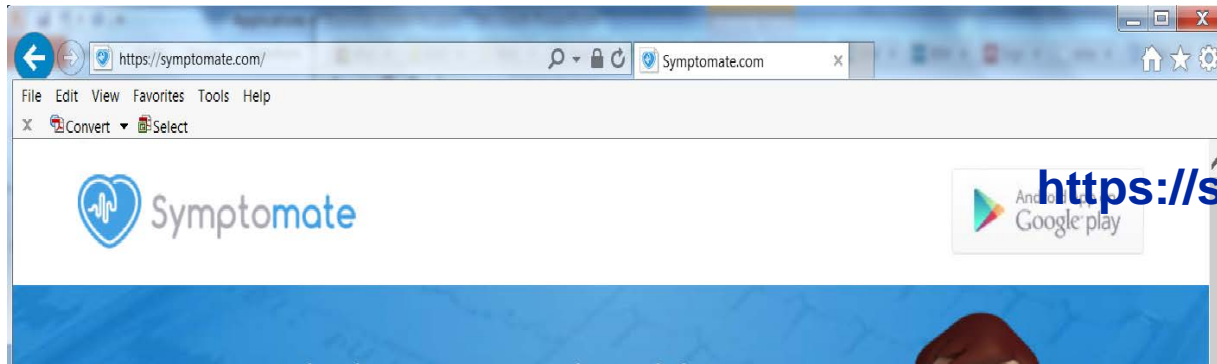
**$10^{82}$  jest szacowaną liczbą atomów w obserwowalnym wszechświecie**

[Przytula et al.] 2.127 zmiennych; 12.351 parametrów numerycznych (zamiast  $2^{2127}$ !)

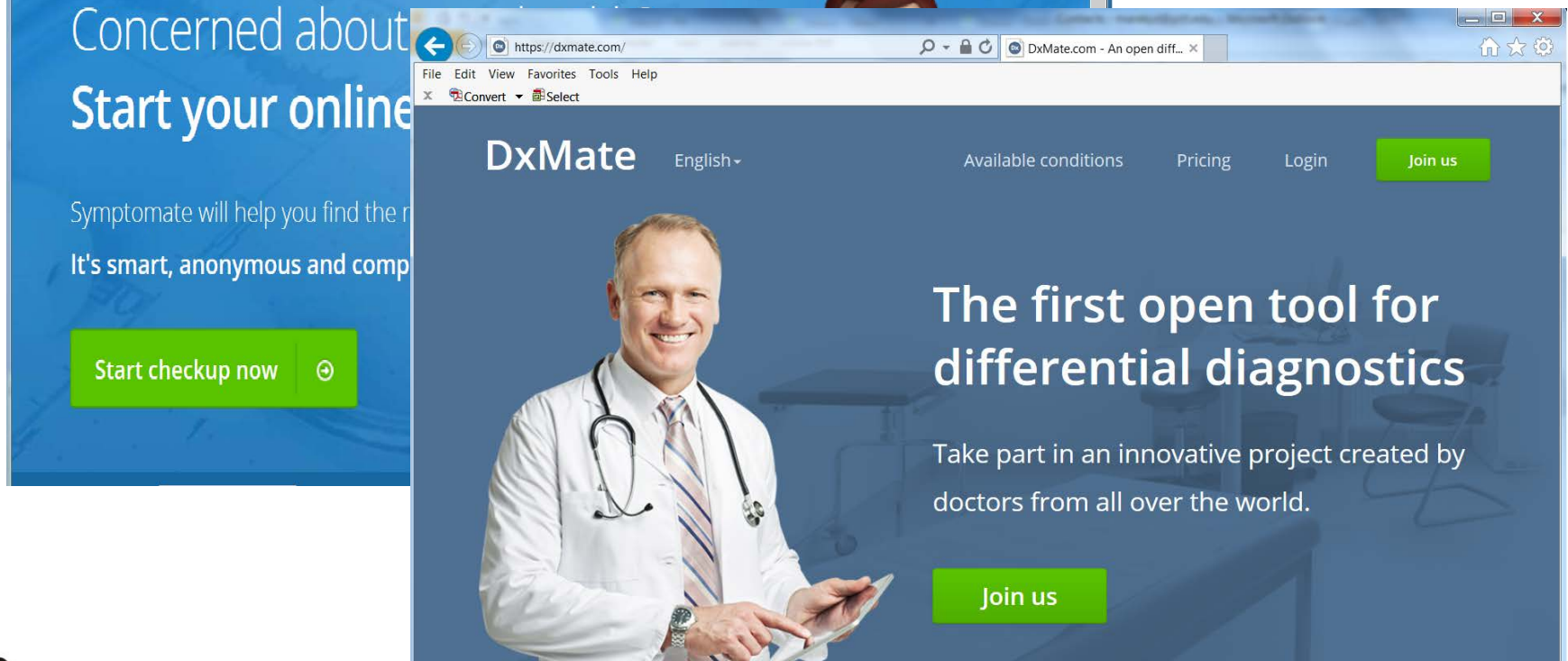


# Dr. Medi: Inteligentny konsultant medyczny

- Wstęp: statystyka bayesowska
- Przyczynowość i prawdopodobieństwo
- Uczenie przyczynowości z danych
- Przykład
- Demo
- Uwagi końcowe



<https://symptomate.com/pl/>



# Przyczynowość i prawdopodobieństwo

## Dlaczego jesteśmy zainteresowani przyczynowością?

**Powód #1:** Przyczynowość pozwala nam przewidywać skutki manipulacji.

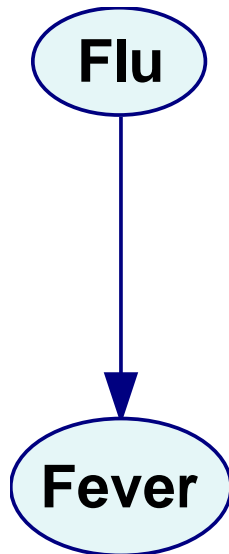
**Powód #2:** Ludzie (w tym eksperci) myślą w kategoriach przyczyn i skutków, więc łatwiej jest budować modele przyczynowe.

**Biorąc pod uwagę (1), czy (2) jest zaskakujące?**

# Przyczynowość i prawdopodobieństwo

Przyczynowość i prawdopodobieństwo są ze sobą ściśle powiązane, a ich związek powinien być jasno określony w statystyce.

Zależność probabilistyczna jest uważana za niezbędny warunek ustalenia związku przyczynowego (czy jest wystarczający?).



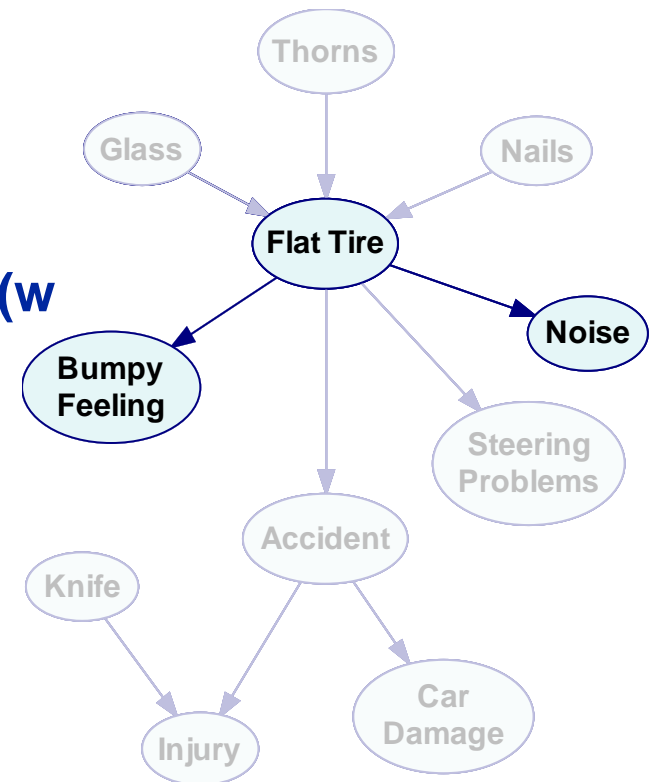
Grypa i gorączka są ze sobą powiązane, **ponieważ** grypa może wywołać gorączkę. Przyczyna może wywołać skutek, ale nie musi. Związki przyczynowe prowadzą do zależności probabilistycznych (lub korelacji w przypadku liniowym).

## Grafy przyczynowo-skutkowe

Acykliczne wykresy skierowane (stąd brak czasu i dynamicznego rozumowania) przedstawiające migawkę świata w danym momencie. Węzły to zmienne losowe, a łuki to bezpośrednie zależności przyczynowe między nimi.

Związki przyczynowe skutkują *korelacją* (w ogólności *zależności probabilistycznej*).

- szkło na drodze będzie skorelowane z przebitymi oponami
- szkło na drodze będzie skorelowane z hałasem
- „telepanie” pojazdu będzie skorelowane z hałasem



## Przynajmniej warunek Markowa

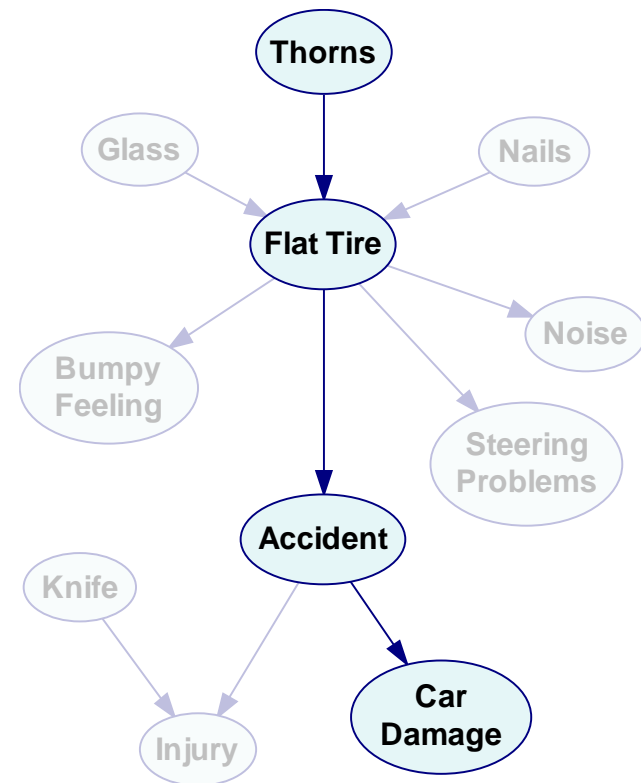
Warunek aksjomatyczny opisujący związek między przyczynowością a prawdopodobieństwem.

**Zmienna w grafie przyczynowym jest niezależna od jej nie-potomków, biorąc pod uwagę jej bezpośrednich poprzedników.**

Aksjomatyczny, ale używany przez prawie wszystkich w praktyce. Jak dotąd nie pokazano żadnych przekonujących przeciwnych przykładów (poza światem kwantowym).

## Warunek Markowa: Implikacje

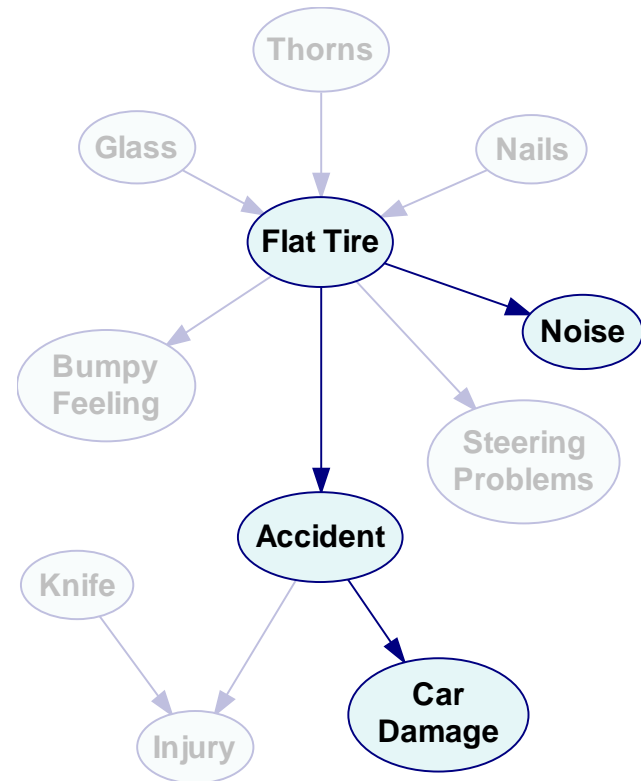
**Zmienne A i B są od siebie zależne, jeśli istnieje skierowana aktywna ścieżka z A do B lub z B do A:**  
**Ciernie na drodze są skorelowane z uszkodzeniem samochodu, ponieważ istnieje ukierunkowana ścieżka od cierni do uszkodzenia samochodu.**



## Warunek Markowa: Implikacje

Zmienne A i B są od siebie zależne, jeśli istnieje C takie, że istnieje skierowana aktywna ścieżka z C do A i istnieje skierowana aktywna ścieżka z C do B:

Uszkodzenie samochodu jest skorelowane z hałasem, ponieważ od przebicia opony przebiega skierowana ścieżka do obu (przebite opony jest wspólną przyczyną obu).

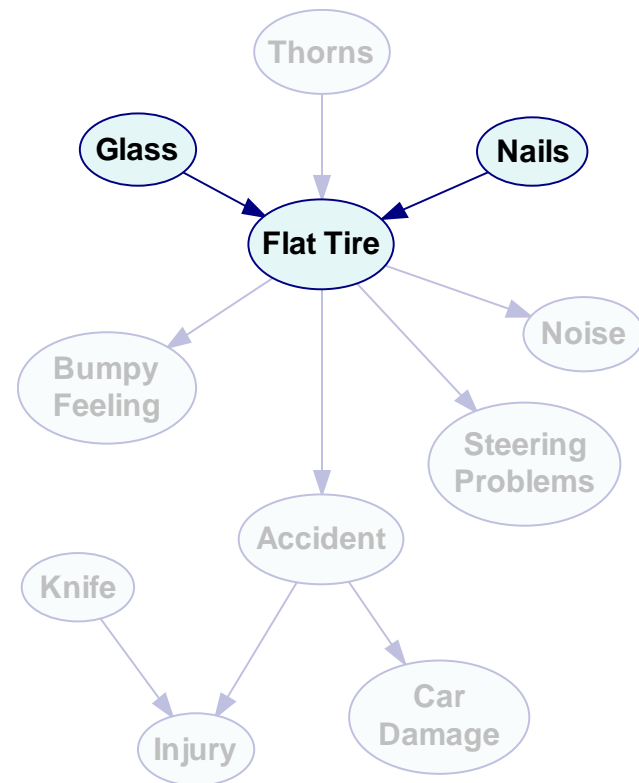




## Warunek Markowa: Implikacje

Zmienne A i B są od siebie zależne, jeśli istnieje D takie, że D jest obserwowane (warunkowane) i istnieje C takie, że A jest zależne od C i istnieje skierowana aktywna ścieżka z C do D i istnieje E takie że B jest zależne od E i istnieje skierowana aktywna ścieżka od E do D:

Gwoździe na drodze są skorelowane ze szkłem na drodze przy przebiciu opony, ponieważ istnieje skierowana ścieżka od szkła na drodze do przebitej opony i od gwoździ na drodze do przebicia opony i przebicia opony (uwarunkowana).



## Warunek Markowa : Podsumowanie konsekwencji

Zmienne A i B są probabilistycznie zależne, jeśli:

- istnieje skierowana aktywna ścieżka z A do B lub istnieje skierowana aktywna ścieżka z B do A
- istnieje C takie, że istnieje skierowana aktywna ścieżka z C do A i istnieje skierowana aktywna ścieżka z C do B
- istnieje D takie, że D jest obserwowane (uwarunkowane) i istnieje C takie, że A jest zależne od C i istnieje skierowana aktywna ścieżka z C do D i istnieje E takie, że B jest zależne od E i istnieje skierowana aktywna ścieżka od E do D.

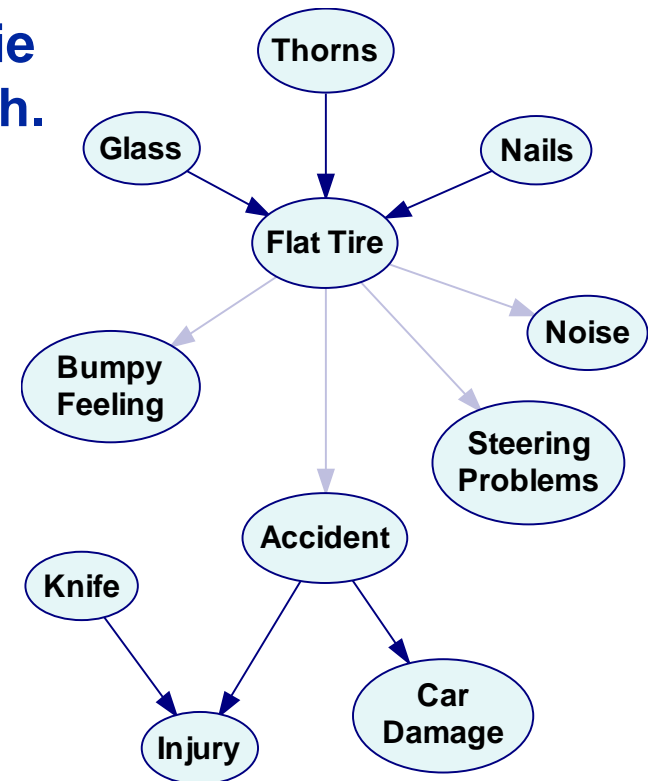
## Warunek Markowa: niezależność warunkowa

Kiedy już poznamy wszystkie bezpośrednie przyczyny zdarzenia E, przyczyny i inne skutki tych przyczyn nie mówią nic nowego o E i jego następach.

(znany również jako „odekranowanie”)

Np.,

- Szkło i kolce na jezdni są niezależne od hałasu, wrażeń wyboistości i problemów z kierowaniem spowodowanych przebicciem opony.
- Hałas, wyboistość i problemy z kierowaniem stają się niezależne od przebiccia opony.

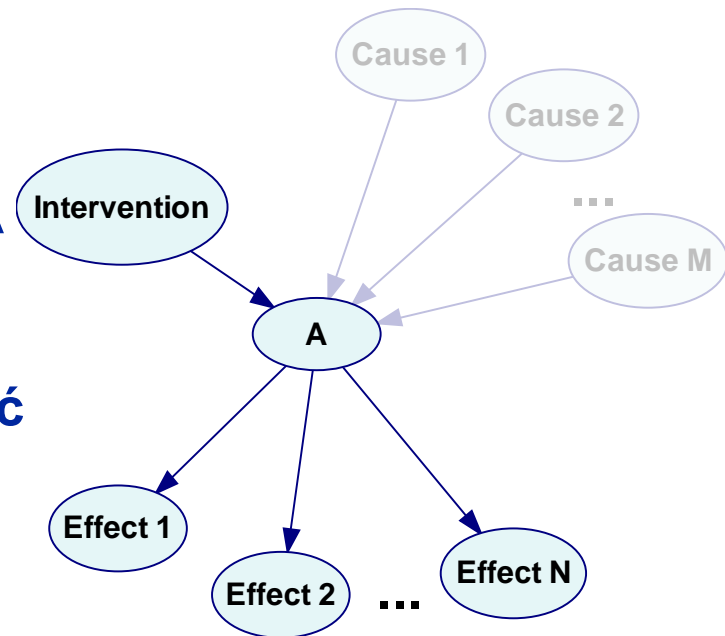


# Interwencja

**Twierdzenie o manipulacji [Spirtes, Glymour & Scheines 1993]:**

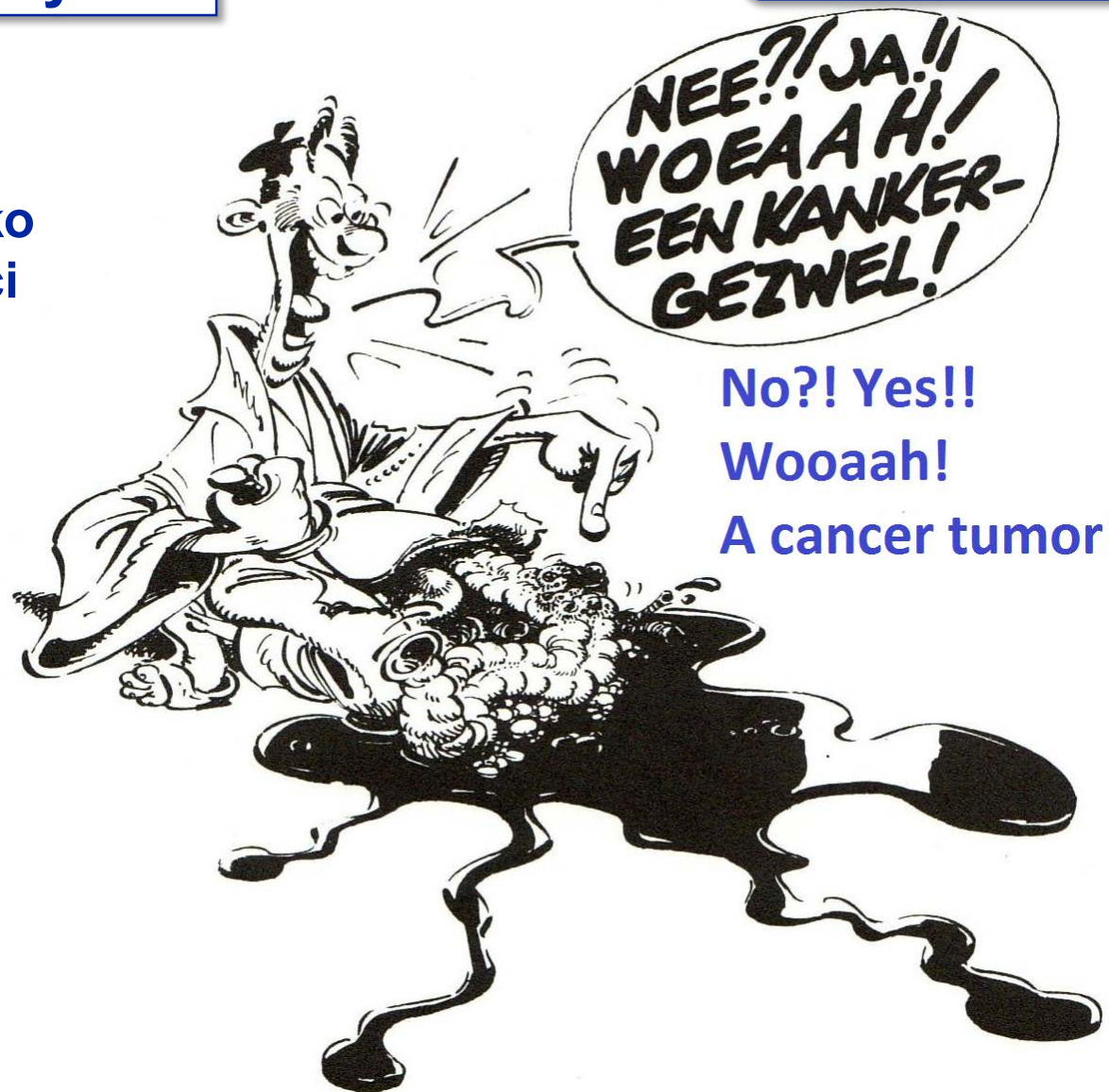
**Przy zewnętrznej interwencji na zmienną A w grafie przyczynowym, możemy wyprowadzić łączny rozkład prawdopodobieństwa w całym grafie modyfikując jedynie warunkowy rozkład prawdopodobieństwa A.**

**Jeśli ta interwencja jest wystarczająco silna, aby ustawić A na określoną wartość, możemy postrzegać tę interwencję jako jedyną przyczynę A i odzwierciedlić to usuwając wszystkie krawędzie, które wchodzą do A. Nic innego w grafie nie wymaga modyfikacji.**



## Interwencja: Przykład

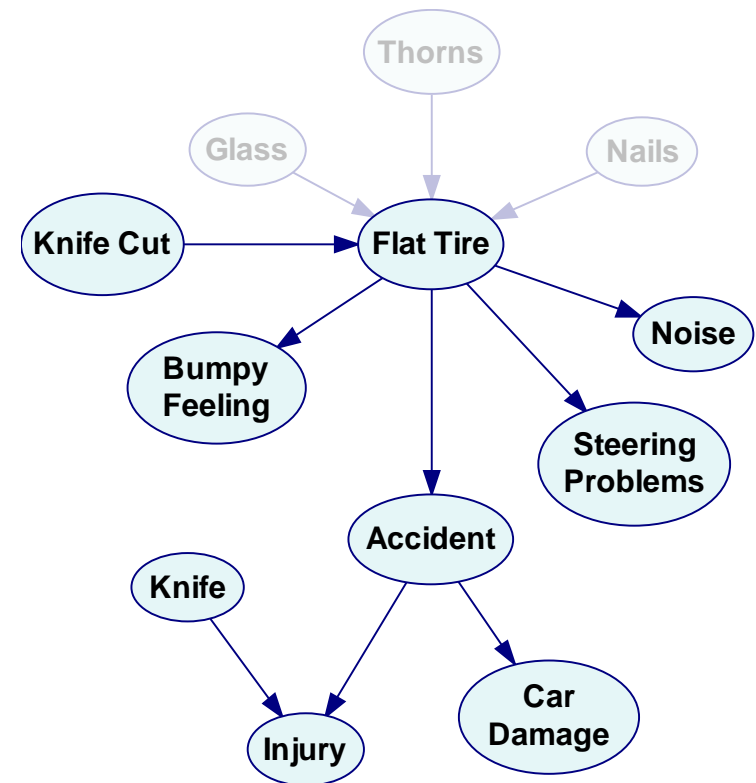
Samobójstwo  
eliminuje raka jako  
przyczynę śmierci  
tego odważnego  
samuraja.



No?! Yes!!  
Wooaah!  
A cancer tumor!

## Interwencja: Przynalynad

Przebicie opony nożem sprawia, że szkło, kolce, gwoździe i wszystko inne, są nieistotne dla przebicia opony. Nóż jest jedyną przynalyną przebicia opony.



# Experymenty naukowe

Badania empiryczne zwykle dotyczą testowania hipotez przyczynowych.

## Palenie i rak płuca są ze sobą skorelowane.

Czy możemy zmniejszyć zachorowalność na raka płuca poprzez ograniczenie palenia?

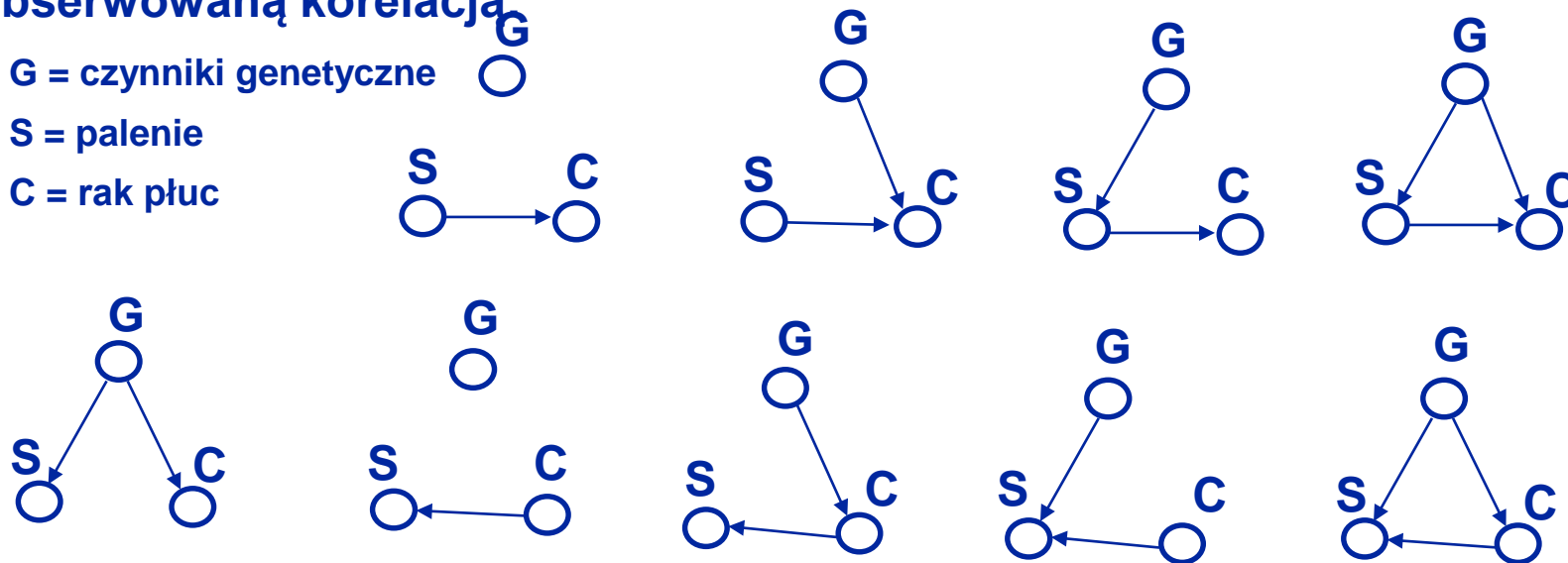
Innymi słowy: czy palenie jest **przyczyną** raka płuca?

Każda z poniższych struktur przyczynowych jest zgodna z obserwowaną korelacją:

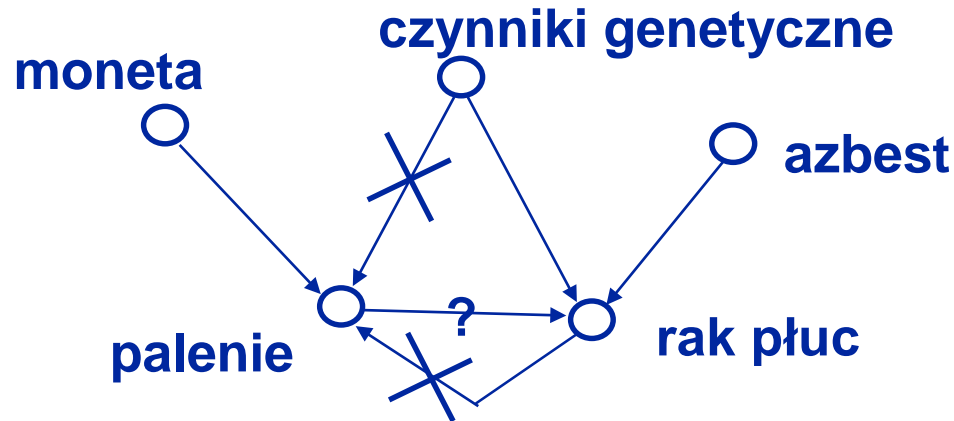
G = czynniki genetyczne

S = palenie

C = rak płuca



## Experymenty naukowe



- W eksperymencie jedyną przyczyną palenia jest moneta.
- Palenie i rak płuc będą zależne od siebie tylko wtedy, gdy istnieje związek przynowocowy pomiędzy paleniem i rakiem płuc.
- Jeśli  $\Pr(C|S) \neq \Pr(C|\sim S)$  palenie jest przyczyną raka płuc.
- Azbest po prostu powoduje zmienność w raku płuc (dodaje szumu do obserwacji).

**Ale czy naprawdę możemy eksperymentować w tej dziedzinie?**



## Nauka poprzez obserwację

**“... Korelacja między paleniem a rakiem płuc oznacza tyle samo, co korelacja między importem jabłek a wzrostem liczby rozwodów ...”**



Sir Ronald A. Fisher, prominentny statystyk, ojciec projektowania eksperymentów



**“... Przypisywanie sobie zakończenia zimnej wojny przez George'a Busha przypomina przypisowanie sobie świtu przez koguta ...”**



Wice-prezydent Al Gore dp wice-prezydenta Dana Quayle podczas ich pierwszej debaty (wice-)prezydenckiej, jesień 1992

# Nauka poprzez obserwację

Wstęp: statystyka bayesowska  
● Przyczynowość i prawdopodobieństwo  
Uczenie przyczynowości z danych  
Przykład  
Demo  
Uwagi końcowe

- Eksperymentowanie nie zawsze jest możliwe.
- Sporo możemy zrobić po prostu obserwując.
- Założenia są kluczowe zarówno w eksperymentowaniu, jak i obserwacji, chociaż w tym drugim przypadku są zwykle silniejsze.
- Nowe metody w odkrywaniu przyczyn: Wyciskanie informacji z danych do granic.

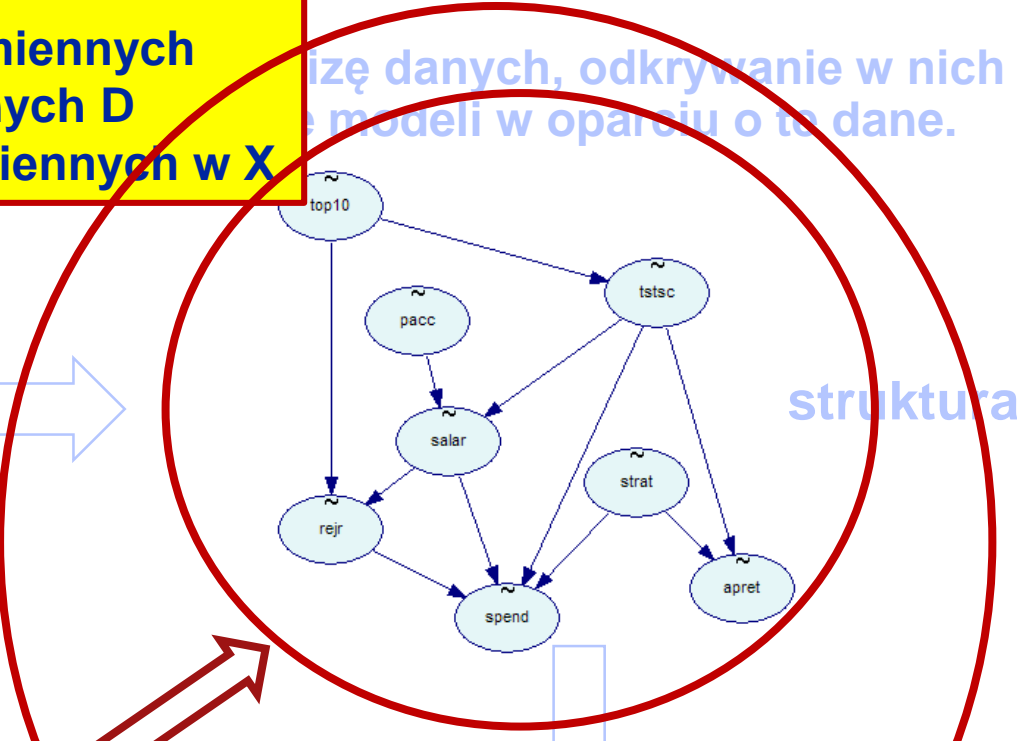
# Uczenie przyczynowości z danych

- Wstęp: statystyka bayesowska
- Przyczynowość i prawdopodobieństwo
- Uczenie przyczynowości z danych
- Przykład
- Demo
- Uwagi końcowe

# Uczenie sieci bayesowskich z danych

**Problem uczenia:**  
 Biorąc pod uwagę zestaw zmiennych (a.k.a. atrybuty) X i zbiór danych D równoczesnych wartości zmiennych w X

| spend | apret  | top10 | rejir  | tspsc  | pacc   | strat | salar |
|-------|--------|-------|--------|--------|--------|-------|-------|
| 9855  | 52.5   | 15    | 29.474 | 65.063 | 36.887 | 12    | 60800 |
| 10527 | 84.25  | 30    | 28.888 | 54.868 | 38.97  | 12.8  | 63900 |
| 7904  | 37.75  | 26    | 25.853 | 60.75  | 41.985 | 20.3  | 57800 |
| 6601  | 57     | 23    | 11.296 | 67.188 | 40.289 | 17    | 51200 |
| 7251  | 62     | 17    | 22.635 | 56.25  | 46.78  | 18.1  | 48000 |
| 6967  | 66.75  | 40    | 9.718  | 65.625 | 53.103 | 18    | 57700 |
| 8489  | 70.333 | 20    | 15.444 | 59.875 | 50.46  | 13.5  | 44000 |
| 9554  | 85.25  | 79    | 44.225 | 74.688 | 40.137 | 17.1  | 70100 |
| 15287 | 65.25  | 42    | 26.913 | 70.75  | 28.276 | 14.4  | 71738 |
| 7057  | 55.25  | 17    | 24.379 | 59.063 | 44.251 | 21.2  | 58200 |
| 16848 | 77.75  | 48    | 26.69  | 75.938 | 27.187 | 9.2   | 63000 |
| 18211 | 91     | 87    | 76.681 | 80.625 | 51.164 | 12.8  | 74400 |
| 21561 | 69.25  | 58    | 44.702 | 76.25  | 26.689 | 9.2   | 75400 |
| 20667 | 65     | 68    | 22.995 | 75.625 | 28.038 | 11    | 66200 |
| 10684 | 61.75  | 26    | 8.774  | 66     | 33.99  | 9.5   | 52900 |
| 11738 | 74.25  | 32    | 25.449 | 66.875 | 27.701 | 12    | 63400 |
| 10107 | 74     | 43    | 11.315 | 71     | 29.096 | 16.2  | 66200 |
| 7817  | 65.75  | 36    | 33.709 | 64.25  | 52.548 | 17.7  | 54600 |
| 7050  | 26     | 11    | 0      | 55.313 | 55.651 | 18.8  | 59500 |
| 9082  | 83.5   | 73    | 64.668 | 77.375 | 43.185 | 13.6  | 66700 |
| 11706 | 60     | 56    | 16.937 | 73.75  | 39.479 | 12.7  | 62100 |



Uzyskaj **wgląd** w związki przyczynowe między zmiennymi w X (w celu zrozumienia i przewidzenia skutków manipulacji)

|         |     |
|---------|-----|
| Success | 0.2 |
| Failure | 0.8 |

|          |         |         |
|----------|---------|---------|
| Success  | Success | Failure |
| Good     | 0.4     | 0.1     |
| Moderate | 0.4     | 0.3     |
| Poor     |         |         |

**Naucz się rozkładu łącznego prawdopodobieństwa zmiennych w X**



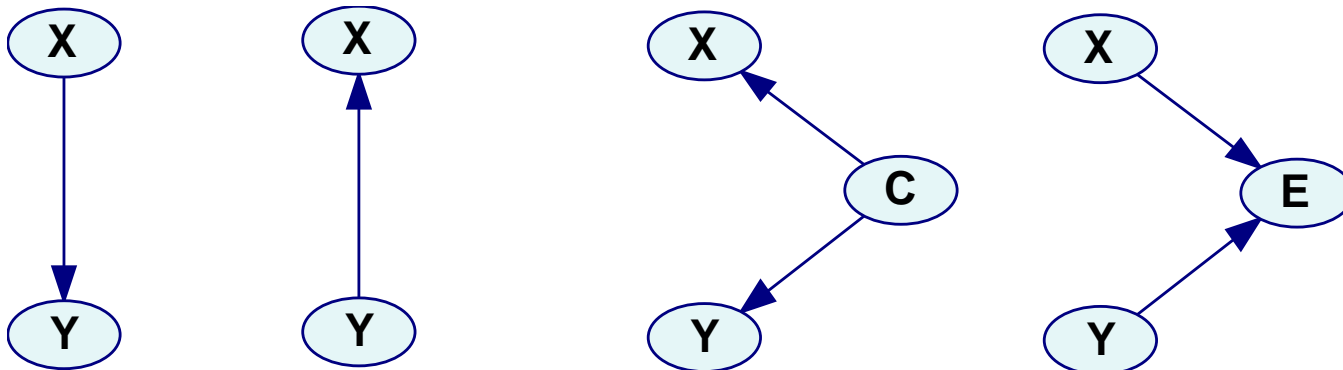
## Korelacja nie oznacza związku przyczynowego

### “Korelacja nie oznacza związku przyczynowego”

Prawda, tylko prawda, ale nie cała prawda.

Prawda tylko w ograniczonych sytuacjach (np. dwie zmienne) i zazwyczaj nadużywana przez autorów podręczników do statystyki 😊.

Jeśli  $x$  i  $y$  są zależne, możemy rzeczywiście uprościć obraz przyczynowy do czterech przypadków:

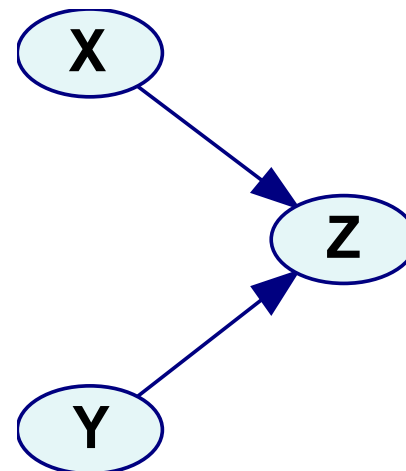


## Korelacja nie oznacza związku przyczynowego

Niekoniecznie prawda w przypadku trzech zmiennych:

x i z są zależne  
y i z są zależne  
x i y są niezależne  
x i y są zależne pod warunkiem z

**Możemy określić  
przyczynowość!**



# Podstawy odkrywania przyczynowości

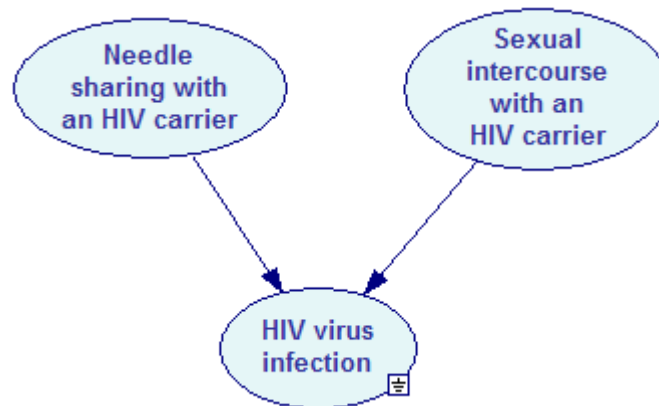
- Warunek Markowa:  
 $d$ -separacja  $\Rightarrow$  niezależność w danych.
- Warunek wierności:  
 $d$ -separacja  $\Leftarrow$  niezależność w danych.

Graf przyczynowo-  
skutkowy określa, co  
jest niezależne.

Wszystkie niezależności w  
danych są strukturalne, tj.  
są konsekwencją warunku  
Markowa.

## Naruszenia warunku wierności

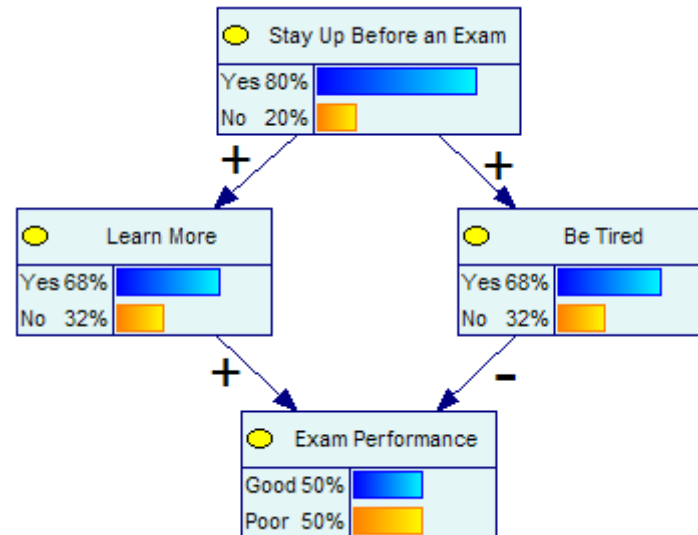
**Założenie wierności jest bardziej kontrowersyjne.  
Podczas gdy każdy naukowiec używa go w  
praktyce, niekoniecznie jest spełnione**



**Biorąc pod uwagę, że nie doszło do zakażenia wirusem HIV, dzielenie się igłami jest niezależne od lekkomyślnego współżycia.**



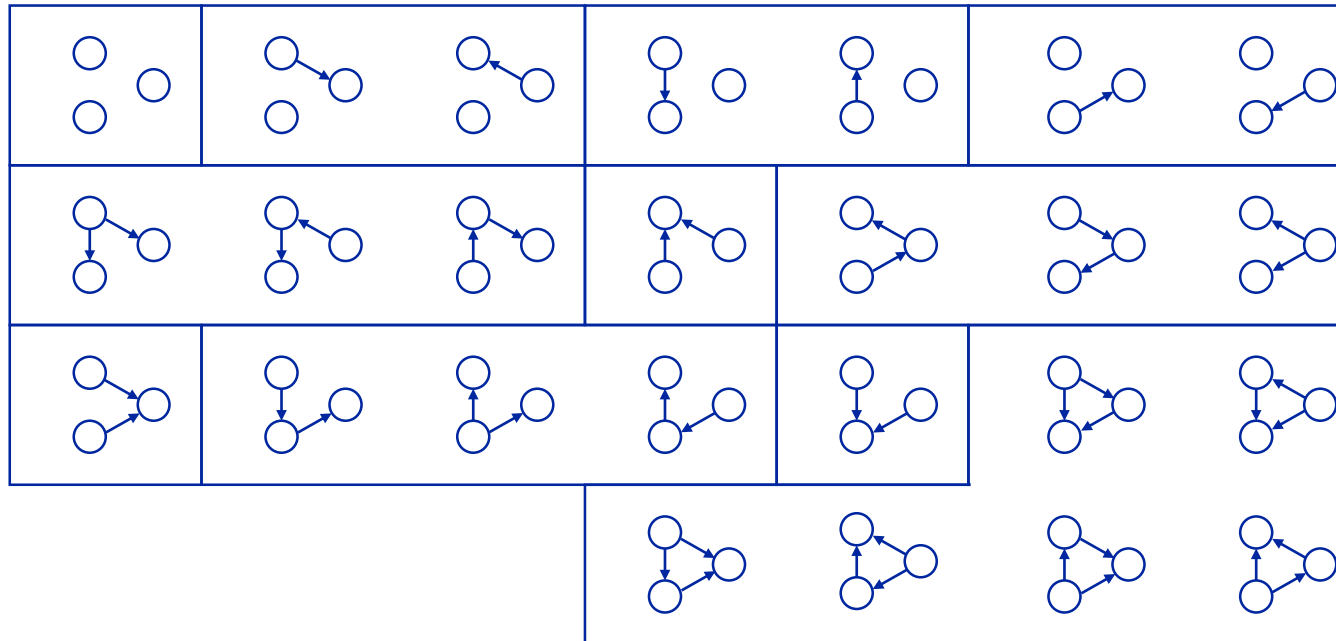
# Naruszenia warunku wierności



**Wpływ siedzenia po nocy przed egzaminem na wynik egzaminu może być zerowy: zmęczenie może niwelować efekt większej wiedzy.  
Ale czy jest to prawdopodobne?**

# Szukanie grafów przyczynowo-skutkowych

Wszystkie możliwe grafy ...



... można podzielić na klasy równoważności

# Twierdzenia użyteczne w szukaniu

## Twierdzenie 1 (szkielet)

Nie ma krawędzi między  $X$  i  $Y$  wtedy i tylko wtedy, gdy  $X$  i  $Y$  są niezależne, warunkując podzbiorem (w tym podzbiorem pustym) innych zmiennych.

## Twierdzenie 2 (struktury $v$ )

Jeśli  $X—Y—Z$ ,  $X$  i  $Z$  nie sąsiadują ze sobą, a  $X$  i  $Z$  są niezależne pod warunkiem pewnym podzbiorem zmiennych  $W$ , to  $X→Y←Z$  wtedy i tylko wtedy, gdy  $W$  nie zawiera  $Y$ .

# Szukanie grafów przyczynowo-skutkowych

1. Znajdź (potencjalnie warunkowe) niezależności w danych.
2. Wnioskuj z tych niezależności, które (klasy) struktur przyczynowych mogły dać początek tym niezależnościom (używając np. algorytmu PC).

## Algorytm PC (szkic)

### Krok 0:

Rozpocznij od nieskierowanego grafu zupełnego.

### Krok 1 (Znajdź połączenia):

Dla każdej pary zmiennych  $\langle X, Y \rangle$  jeśli  $X$  i  $Y$  są niezależne przy warunkowaniu podzbiorem innych zmiennych, usuń krawędź  $X-Y$ .

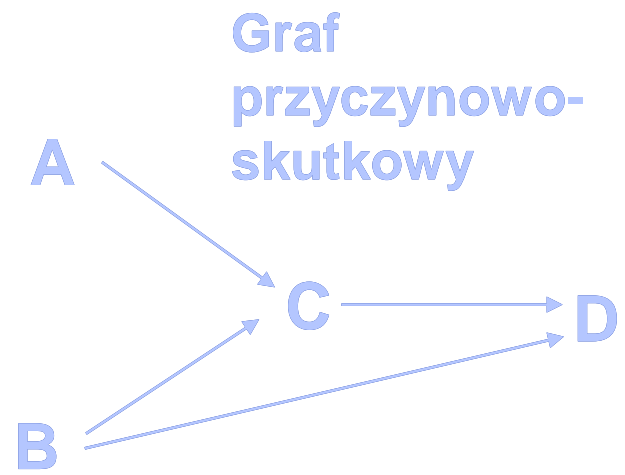
### Krok 2: (Znajdź struktury $v$ ):

Dla każdej trójki  $X-Y-Z$ , jeśli nie ma połączenia bezpośredniego pomiędzy  $X$  i  $Z$ , jeśli  $X$  i  $Z$  są niezależne przy warunkowaniu podzbiorem zmiennych niezawierającym  $Y$ , zorientuj  $X-Y-Z$  jako  $X \rightarrow Y \leftarrow Z$ .

### Krok 3 (Uniknij nowych struktur $v$ i cykli skierowanych):

- Jeśli istnieje  $X \rightarrow Y - Z$ , ale nie ma krawędzi pomiędzy  $X$  i  $Z$ , zorientuj  $Y-Z$  jako  $Y \rightarrow Z$ .
- Jeśli istnieje  $X - Z$  i istnieje w grafie skierowana ścieżka od  $X$  do  $Z$ , zorientuj  $X - Z$  jako  $X \rightarrow Z$ .

# Algorytm PC: Przykład

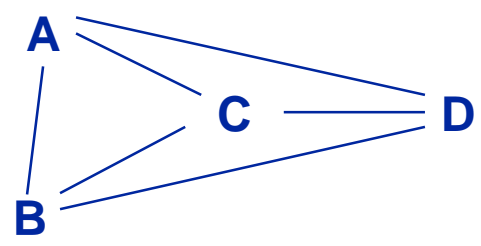


Niezależności, które zauważymy w danych:

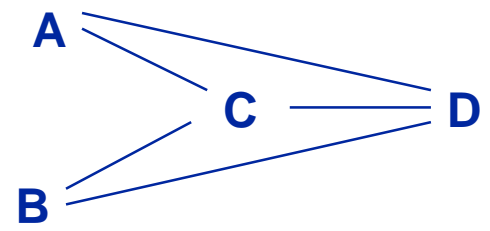
$$A \perp B$$

$$A \perp D \mid B, C$$

(0) Rozpocznij od grafu zupełnego

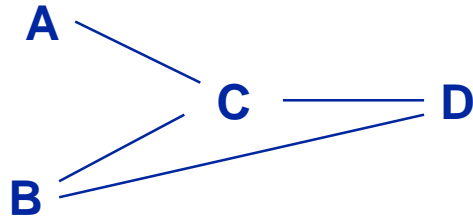


(1) Na podstawie  $A \perp B$ , usuń  $A-B$

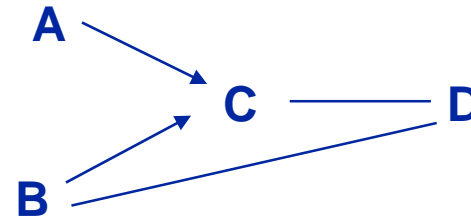


# Algorytm PC: Przykład

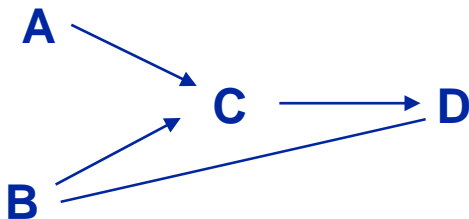
(1) Na podstawie  $A \perp D \mid B, C$ , usuń  $A-D$



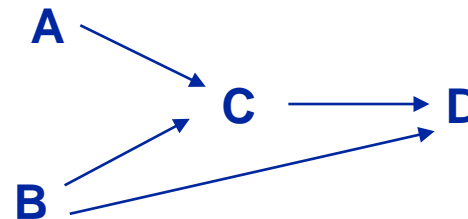
(2) Na podstawie  $A \perp B$ , skieruj  $A-C-B$  jako  $A \rightarrow C \leftarrow B$



(3) Aby uniknąć struktury v ( $A \rightarrow C \leftarrow D$ ), skieruj  $C-D$  jako  $C \rightarrow D$ .



(3) Aby uniknąć cyklu ( $B \rightarrow C \rightarrow D \rightarrow B$ ), zorientuj  $B-D$  jako  $B \rightarrow D$ .



## „Wzorce”: rezultat algorytmu PC

Algorytm PC generuje „wzorzec”, rodzaj wykresu zawierającego skierowane ( $\rightarrow$ ), dwukierunkowe ( $\leftrightarrow$ ) i nieskierowane ( $—$ ) krawędzie, które reprezentują klasę równoważności Markowa modeli

- Skierowana krawędź  $A \rightarrow B$  we „wzorcu” wskazuje, że w każdym grafie w klasie równoważności Markowa istnieje krawędź skierowana  $A \rightarrow B$
- Dwukierunkowa krawędź  $A \leftrightarrow B$  w „wzorcu” wskazuje, że istnieje krawędź między  $A$  i  $B$  na każdym wykresie w klasie równoważności Markowa, chociaż jej kierunku nie można ustalić na podstawie danych
- Niekierowana krawędź  $A — B$  we „wzorcu” wskazuje, że istnieje krawędź między  $A$  i  $B$  na każdym wykresie w klasie równoważności Markowa, chociaż jej kierunku nie można ustalić na podstawie danych; może też istnieć wspólna przyczyna między tymi zmiennymi na każdym wykresie w klasie równoważności Markowa

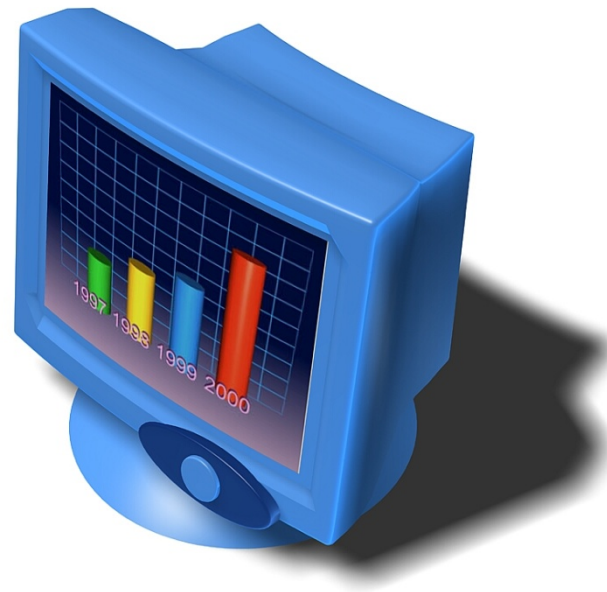


## Przykładowe zastosowanie

- **Utrzymanie studentów na uczelniach w USA.**
- **Duży problem dla amerykańskich uczelni.**
- **Używając tych metod, prawidłowo przewidzieliśmy, że głównym czynnikiem sprawczym niskiego wskaźnika retencji studentów jest jakość przyjętych studentów [Druzdzel & Glymour, 1994]**

# Pozostałość tej sesji

Wstęp: statystyka bayesowska  
Przyczynowość i prawdopodobieństwo  
Uczenie przyczynowości z danych  
Przykład  
● Demo  
Uwagi końcowe



## Uwagi końcowe

- Obserwacja jest ważną metodą naukową.
- Obserwacja często pozwala ograniczyć klasę możliwych struktur przyczynowych, które mogły wygenerować dane.
- Uczenie się sieci bayesowskich / grafów przyczynowych jest bardzo ekscytujące: jest to inny sposób uprawiania nauki.
- Istnieje bogaty asortyment nierozwiązanych problemów związanych z odkrywaniem przyczyn / uczeniem się sieci bayesowskich, zarówno praktycznych, jak i teoretycznych.
- Uczenie maszynowe jest aktywnym obszarem moich badań (wytworem tej pracy jest GeNIe, <https://www.bayesfusion.com/>).

