# Sztuczna Inteligencja i Systemy Doradcze

## Sieci bayessowskie 1

# Niepewnosc

Niech akcja $A_t$ = wyjedź na lotnisko $t$ minut przed odlotem
Czy $A_t$ pozwoli mi zdążyć na czas?

Problemy:

1) informacja częściowa (stan ulic, plany innych kierowców, etc.)
2) niedokładne informacje (raport o korkach)
3) niepewność działania akcji (złapanie gumy, etc.)
4) ogromna złożoność modelowania i przewidywania ruchu

Stąd czysto logiczne podejście albo

1) ryzykuje fałszywość: "$A_{25}$ pozwoli mi zdążyć na czas"

albo 2) prowadzi do wniosków zbyt słabych do podjęcia decyzji:

"$A_{25}$ pozwoli mi zdążyć na czas jeśli nie będzie wypadku na moście i nie będzi padać i nie złapię gumy itd."

($A_{1440}$ mogłoby być uznane że rozsądnie zapewnia, że zdąże na czas, ale nie chcę czekać całą noc na lotnisku ...)

# Metody wnioskowania w niepewnosci

Logika defaultowa lub niemonotoniczna:

Załóż, że samochód nie złapie gumy

Załóż, że $A_{25}$ działa, jeśli nie ma sprzecznych przesłanek

Pytania: Jakie założenia są sensowne? Jak zarządzać sprzecznościami?

Reguły z czynnikiem ryzyka:

$A_{25} \mapsto_{0.3}$ zdąży na czas

$Zraszacz \mapsto_{0.99} MokryTrawnik$

$MokryTrawnik \mapsto_{0.7} Deszcz$

Pytania: Problemy z kombinowaniem, np. czy $Zraszacz$ powoduje $Deszcz$??

Prawdopodobieństwo

Dla dostępnych przesłanek

$A_{25}$ zdąży na czas z prawdopodobieństwem 0.04

Mahaviracarya (IX w.), Cardamo (1565) teoria ryzyka

(Logika rozmyta zarządza *stopniem prawdziwości* NIE niepewnością np.

$MokryTrawnik$ jest prawdą w stopniu 0.2)

# Prawdopodobienstwo

Stwierdzenia prawdopodobne *zbierają* efekt

ograniczenia: niemożność wyliczenia wyjątków, warunków, etc.

braku wiedzy: brak istotnych faktów, warunków początkowych, etc.

Prawdopodobieństwo subiektywne lub bayessowskie:

Prawdopodobieństwa odnoszą stwierdzenia do czyjegoś stanu wiedzy

np. $P(A_{25}|\text{brak zgłoszonych wypadków}) = 0.06$

To *nie* są stwierdzenia o prawdopodobnej tendecji w bieżącej sytuacji (ale mogłyby być wyuczone ze zdobytego doświadczenia lub podobnych sytuacji)

Pradowpodobieństwo zdarzenia zmienia się wraz z nową przesłanką:

np. $P(A_{25}|\text{brak zgłoszonych wypadków, 5-ta rano}) = 0.15$

# Podejmowanie decyzji w niepewnosci

Załóżmy, ze wierzę w następujące zdania:

$$
\begin{aligned}
P(A_{25} \text{ pozwoli zdążyć na czas}| \ldots) &= 0.04 \\
P(A_{90} \text{ pozwoli zdążyć na czas}| \ldots) &= 0.70 \\
P(A_{120} \text{ pozwoli zdążyć na czas}| \ldots) &= 0.95 \\
P(A_{1440} \text{ pozwoli zdążyć na czas}| \ldots) &= 0.9999
\end{aligned}
$$

Którą akcję wybrać?

Zależy od moich preferencji co do spóźnienia, kuchni lotniska, itd.

Teoria użyteczności jest używana do reprezentacji i wnioskowania o preferencjach

Teoria decyzji = teoria użyteczności + teoria prawdopodobieństwa

# Podstawy prawdopodobienstwa

$\Omega$ — *przestrzeń próbek*

np. 6 możliwych wyników rzutu kostką.

$\omega \in \Omega$ jest punktem próbkowym/dopuszczalnym stanem świata/ zdarzeniem atomowym

*Przestrzeń prawdopobieństwa* lub *model prawdopodobieństwa* to przestrzeń próbek z przypisaniem $P(\omega)$ dla każdego $\omega \in \Omega$ spełniającego warunki

$0 \leq P(\omega) \leq 1$

$\sum_\omega P(\omega) = 1$

np. $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

*Zdarzenie* $A$ jest podzbiorem $\Omega$

$$P(A) = \sum_{\{\omega \in A\}} P(\omega)$$

Np. $P(\text{rzut kostką} < 4) = 1/6 + 1/6 + 1/6 = 1/2$

# Zmienne losowe

*Zmienna losowa* jest funkcją z przestrzeni próbek w pewien zbiór wartości,

np. rzeczywistych lub boolowskich

np. $Odd(1) = true$.

$P$ indukuje *rozkład prawdopodbieństwa* dla dowolnej zm. los. $X$:

$$P(X = x_i) = \Sigma_{\{\omega : X(\omega) = x_i\}} P(\omega)$$

np. $P(Odd = true) = 1/6 + 1/6 + 1/6 = 1/2$

# Zdania

Zdania reprezentują pewne zdarzenia (podzbiory przestrzeni próbek)
w których są prawdziwe

Dla danych zmiennych boolowskich $A$ i $B$:

zdarzenie $a$ = zbiór punktów próbkowych gdzie $A(\omega) = true$

zdarzenie $\neg a$ = zbiór punktów próbkowych gdzie $A(\omega) = false$

zdarzenie $a \wedge b$ = zbiór punktów gdzie $A(\omega) = true$ i $B(\omega) = true$

Często w zastosowaniach SI, punkty próbkowe są *definiowane* przez wartości zbioru zmiennych losowych, tzn. przestrzeń próbek jest produktem kartezjańskim zbioru wartości tych zmiennych

Dla zmiennych boolowskich, punkt próbkowy = model rachunku zdań

np. $A = true$, $B = false$, lub $a \wedge \neg b$.

Zdanie = alternatywa zdarzeń atomowych, w których to zdanie jest prawdą

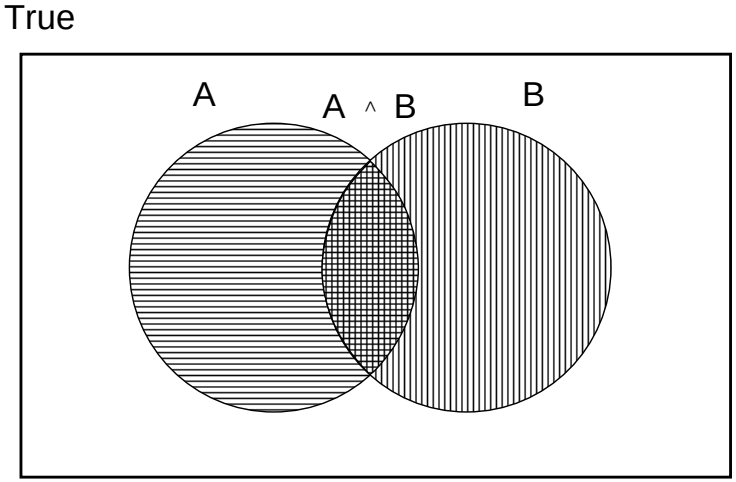
np., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

# Dlaczego uzywac prawdopodobienstwa?

Defincje implikują, że pewne logicznie powiązane zdarzenia muszą mieć powiązane prawdopodobieństwa

Np. $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



# Skladnia zdan

Boolowskie zmienne losowe

np. $Cavity$ (czy jestem osłabiony?)

Dyskretne zmienne losowe (*skończone* lub *nieskończone*)

np. $Weather$ ma jedną wartość z $\langle sunny, rain, cloudy, snow \rangle$

$Weather = rain$ jest zdaniem

Wartości muszą być kompletne i wzajemnie się wykluczać

Ciągłe zmienne losowe (*ograniczone* lub *nieograniczone*)

np. $Temp = 21.6$; można także $Temp < 22.0$.

Dowolne kombinacje boolowskie prostych zdań

# Prawdopodobienstwo bezwarunkowe

Bezwarunkowe prawdopodobieństwo zdań

np. $P(Cavity = true) = 0.1$ i $P(Weather = sunny) = 0.72$

odpowiada przekonaniom przed dostarczeniem jakiejkolwiek (nowej) przesłanki

Rozkład prawdopodobieństwa daje wartości dla wszystkich przypisań:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (*znormalizowana*: sumuje się do 1)

Łączny rozkład prawdopodobieństwa dla zbioru zm. los. daje prawdopodobieństwa każdego zdarzenia atomowego na tych zm. los. (tzn. każdy punkt próbkowy)

$\mathbf{P}(Weather, Cavity)$ = macierz wartości $4 \times 2$:

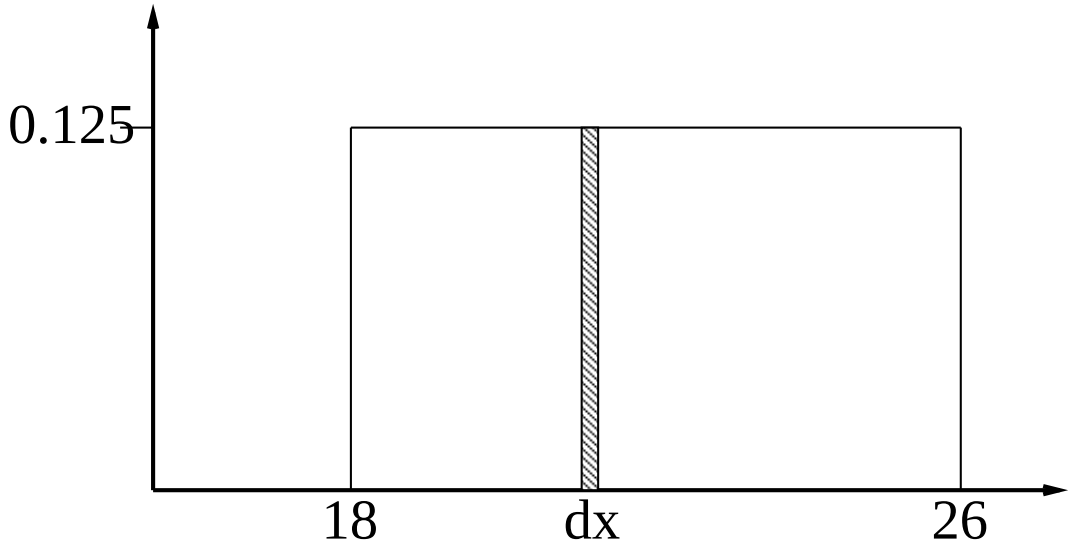
| $Weather =$ | $sunny$ | $rain$ | $cloudy$ | $snow$ |
|---|---|---|---|---|
| $Cavity = true$ | 0.144 | 0.02 | 0.016 | 0.02 |
| $Cavity = false$ | 0.576 | 0.08 | 0.064 | 0.08 |

*Każde pytanie o dziedzinę może być odpowiedziane przez łączny rozkład ponieważ każde zdarzenie jest sumą punktów próbkowych*

# Prawdopodobienstwo dla zmiennych ciaglych

Wyraża rozkład jako parametryzowaną funkcję wartości zmiennej:

$P(X = x) = U[18, 26](x)$ = jedorodny rozkład pomiędzy 18 i 26



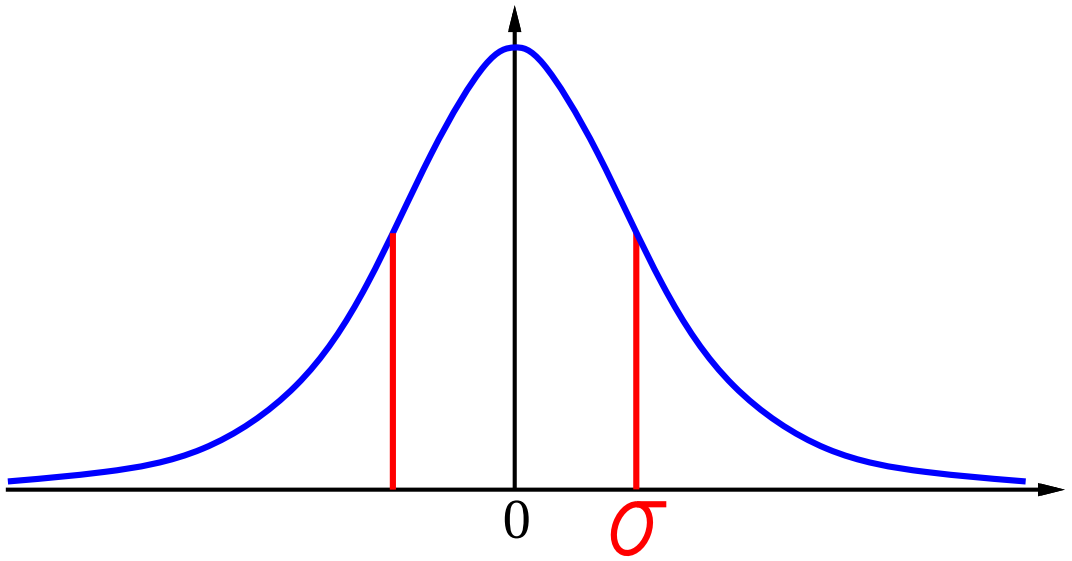

$P$ jest tutaj *gęstością*; całkuje się do 1.

$P(X = 20.5) = 0.125$ oznacza dokładnie

$$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$$

# Rozklad normalny (gaussowski)

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$



# Prawdopodobienstwo warunkowe

Prawdopodobieństwo warunkowe lub a posteriori

np. $P(cavity|toothache) = 0.8$

tzn. zakładając, że $toothache$ to to, o czym wiem

NIE "jeśli $toothache$ to 80% szans na $cavity$"

Notacja rozkładów warunkowych:

$\mathbf{P}(Cavity|Toothache)$ = 2-elementowy wektor 2-elementowych wektorów

Jeśli wiemy więcej, np. $cavity$ też jest dane, wtedy mamy

$P(cavity|toothache, cavity) = 1$

Uwaga: mniej specyficzne przekonania *pozostają prawdziwe*

po dojściu nowych przesłanek, ale nie zawsze są *użyteczne*

Nowe przesłanki mogą być nieistotne, umożliwiając upraszczanie, np.

$P(cavity|toothache, 49ersWin) = P(cavity|toothache) = 0.8$

Ten rodzaj wnioskowania, uwarunkowany wiedzą dziedzinową, jest kluczowy

# Prawdopodobienstwo warunkowe

Definicja prawdopobieństwa warunkowego:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Reguła produkcji daje sformułowanie alternatywne:

$$P(a \wedge b) = P(a|b)P(b) = p(b|a)P(a)$$

Ogólna wersja zachodzi dla całych rozkładów, np.

$$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$$

(jako zbiór $4 \times 2$ równań, *nie* mnożenie macierzy)

Reguła łańcuchowa otrzymywana przez kolejne zastosowania reguły produkcji:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n_1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})\end{aligned}$$

# Wnioskowanie przez wyliczanie

Zacznij od rozkładu łącznego:

| | *toothache* | | ¬ *toothache* | |
|---|---|---|---|---|
| | *catch* | ¬ *catch* | *catch* | ¬ *catch* |
| *cavity* | **.108** | **.012** | **.072** | **.008** |
| ¬ *cavity* | **.016** | **.064** | **.144** | **.576** |

Dla dowolnego zdania $\phi$, sumuj zdarzenia atomowe, w których to zdanie jest prawdziwe:

$$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$$

# Wnioskowanie przez wyliczanie

Zacznij od rozkładu łącznego:

| | *toothache* | | ¬ *toothache* | |
|---|---|---|---|---|
| | *catch* | ¬ *catch* | *catch* | ¬ *catch* |
| *cavity* | **.108** | **.012** | **.072** | **.008** |
| ¬ *cavity* | **.016** | **.064** | **.144** | **.576** |

Dla dowolnego zdania $\phi$, sumuj zdarzenia atomowe, w których to zdanie jest prawdziwe:

$$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$$

$$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

# Wnioskowanie przez wyliczanie

Zacznij od rozkładu łącznego:

| | *toothache* | | ¬ *toothache* | |
|---|---|---|---|---|
| | *catch* | ¬ *catch* | *catch* | ¬ *catch* |
| *cavity* | **.108** | **.012** | **.072** | **.008** |
| ¬ *cavity* | **.016** | **.064** | **.144** | **.576** |

Dla dowolnego zdania $\phi$, sumuj zdarzenia atomowe, w których to zdanie jest prawdziwe:

$$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$$

$$P(cavity \vee toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

# Wnioskowanie przez wyliczanie

Zacznij od rozkładu łącznego:

| | *toothache* | | ¬ *toothache* | |
|---|---|---|---|---|
| | *catch* | ¬ *catch* | *catch* | ¬ *catch* |
| *cavity* | **.108** | **.012** | **.072** | **.008** |
| ¬ *cavity* | **.016** | **.064** | **.144** | **.576** |

Można również policzyć prawdopodobieństwa warunkowe:

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

# Wnioskowanie przez wyliczanie: normalizacja

| | *toothache* | | ¬ *toothache* | |
|---|---|---|---|---|
| | *catch* | ¬ *catch* | *catch* | ¬ *catch* |
| *cavity* | **.108** | **.012** | **.072** | **.008** |
| ¬ *cavity* | **.016** | **.064** | **.144** | **.576** |

Mianownik można traktować jako *stała normalizacji* $\alpha$

$$\begin{aligned}
&\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache) \\
&= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\
&= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] \\
&= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle
\end{aligned}$$

# Wnioskowanie przez wyliczanie

Zazwyczaj interesuje nas

rozkład warunkowy zadanych zmiennych $\mathbf{Y}$

przy danych specyficznych wartościach $\mathbf{e}$ dla zmiennych-przesłanek $\mathbf{E}$

Zmienne ukryte $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Ogólny pomysł: ustalamy zmienne-przesłanki i sumujemy prawdopodobieństwa po wartościach zmiennych ukrytych:

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e}) = \alpha\Sigma_{\mathbf{h}}\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e},\mathbf{H}=\mathbf{h})$$

Wyrażenia w sumowania są wartościami łącznego rozkładu ponieważ $\mathbf{Y}$, $\mathbf{E}$ i $\mathbf{H}$ razem wyczerpują cały zbiór zmiennych losowych

Problemy:

1) Złożoność czasowa $O(d^n)$ gdzie $d$ jest maks. liczbą wartości zmiennej

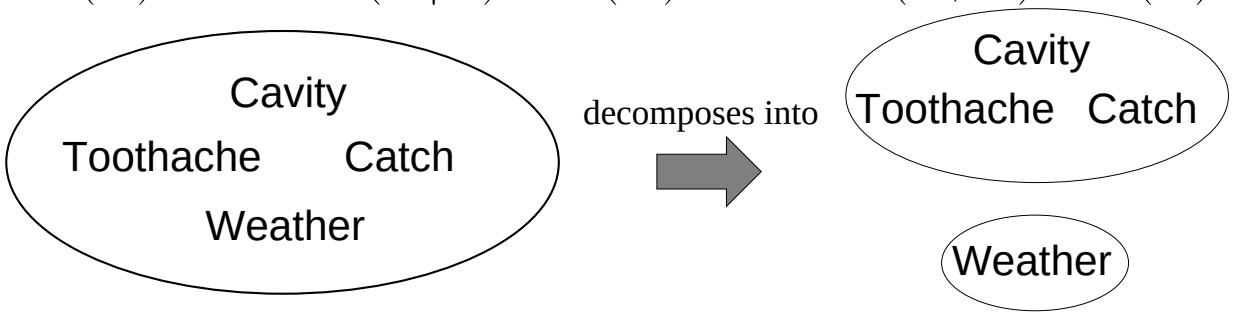
2) Złożoność pamięciowa $O(d^n)$, żeby pamiętać łączny rozkład

3) Jak zbudować słownik wartości prawdopodbieństw dla $O(d^n)$ punktów próbkowych???

# Niezaleznosc

$A$ i $B$ są niezależne wtw

$\mathbf{P}(A|B) = \mathbf{P}(A)$ lub $\mathbf{P}(B|A) = \mathbf{P}(B)$ lub $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$




$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

32 wartości prawdopodbieństw zredukowane do 12; dla $n$ niezależnych rzutów monetą $2^n \rightarrow n$

Pełna niezależność zmiennych jest bardzo efektywna, ale bardzo rzadka

# Niezaleznosc warunkowa

$\mathbf{P}(Toothache, Cavity, Catch)$ wymaga $2^3 - 1 = 7$ niezależnych wartości

Jeśli mam osłabienie, prawdopodbieństwo, że złapię wtedy przeziębienie jest niezależne od tego, czy mam ból zęba:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

Ta sama niezależność pozostaje, jeśli nie mam osłabienia:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

$Catch$ jest *warunkowo niezależne* od $Toothache$ przy danym $Cavity$:

$\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$

Równoważne zdania:

$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$

$\mathbf{P}(Toothache, Catch|Cavity) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$

# Niezaleznosc warunkowa

Używając pełnego łącznego rozkładu i reguły łańcuchowej:

$$\begin{aligned}
&\mathbf{P}(Toothache, Catch, Cavity) \\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity) \\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity) \\
&= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

Tzn. $2 + 2 + 1 = 5$ niezależnych wartości (równania 1 i 2 usuwają 2)

*W większości przypadków użycie prawdopodobieństwa warunkowego redukuje rozmiar reprezentacji łącznego rozkładu z wykładniczego od $n$ do linowego od $n$.*

*Niezależność warunkowa jest najbardziej podstawową i efektywną formą wiedzy o niepewnym środowisku.*

# Regula Bayessa

Reguła produkcytjna $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$$\Rightarrow \text{ reguła Bayessa } P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

lub dla rozkładów

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha\mathbf{P}(X|Y)\mathbf{P}(Y)$$

Użyteczne przy szacowaniu prawdopodobieństwa diagnostycznego na podstawie prawdopodobieństwa przyczynowego:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

Np. $M$ dolegliwość meningitis, $S$ sztywnienie szyji:

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

# Regula Bayessa i niezaleznosc warunkowa

$$\begin{aligned}
&\mathbf{P}(Cavity|toothache \wedge catch) \\
&= \alpha\,\mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity) \\
&= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

Model wnioskowania *naiwny Bayessowski* (zakłada niezależność obserwacji):

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\prod_i \mathbf{P}(Effect_i|Cause)$$


Całkowita liczba parametrów *liniowa* od $n$

# Sieci bayessowskie

Prosta, grafowa notacja do reprezentacji stwierdzeń o niezależności warunkowej i do zwartej specyfikacji pełnych rozkładów wielu zmiennych losowych

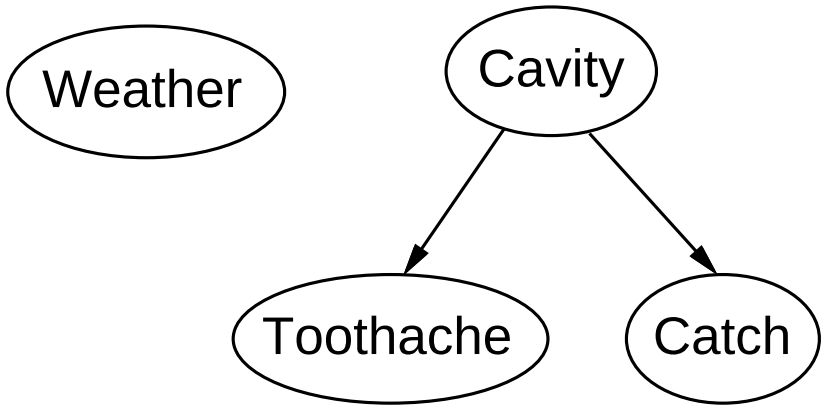
Składnia:

zbiór węzłów, jeden dla każdej zmiennej losowej
skierowany graf acykliczny (strzałka $\approx$ "bezpośrednio wpływa na")
dla każdego węzła rozkład warunkowy na podstawie rodziców:
$\mathbf{P}(X_i|Parents(X_i))$

W najprostszym przypadku rozkład warunkowy reprezentowany jest jako tablica prawdopodobieństwa warunkowego (TPW) dająca rozkład $X_i$
dla każdej kombinacji wartości rodziców

# Sieci bayessowskie: przyklad

Topologia sieci koduje stwierdzenie o warunkowej niezależności:




$Weather$ jest niezależna od innych zmiennnych

$Toothache$ i $Catch$ są warunkowo niezależne przy danym $Cavity$

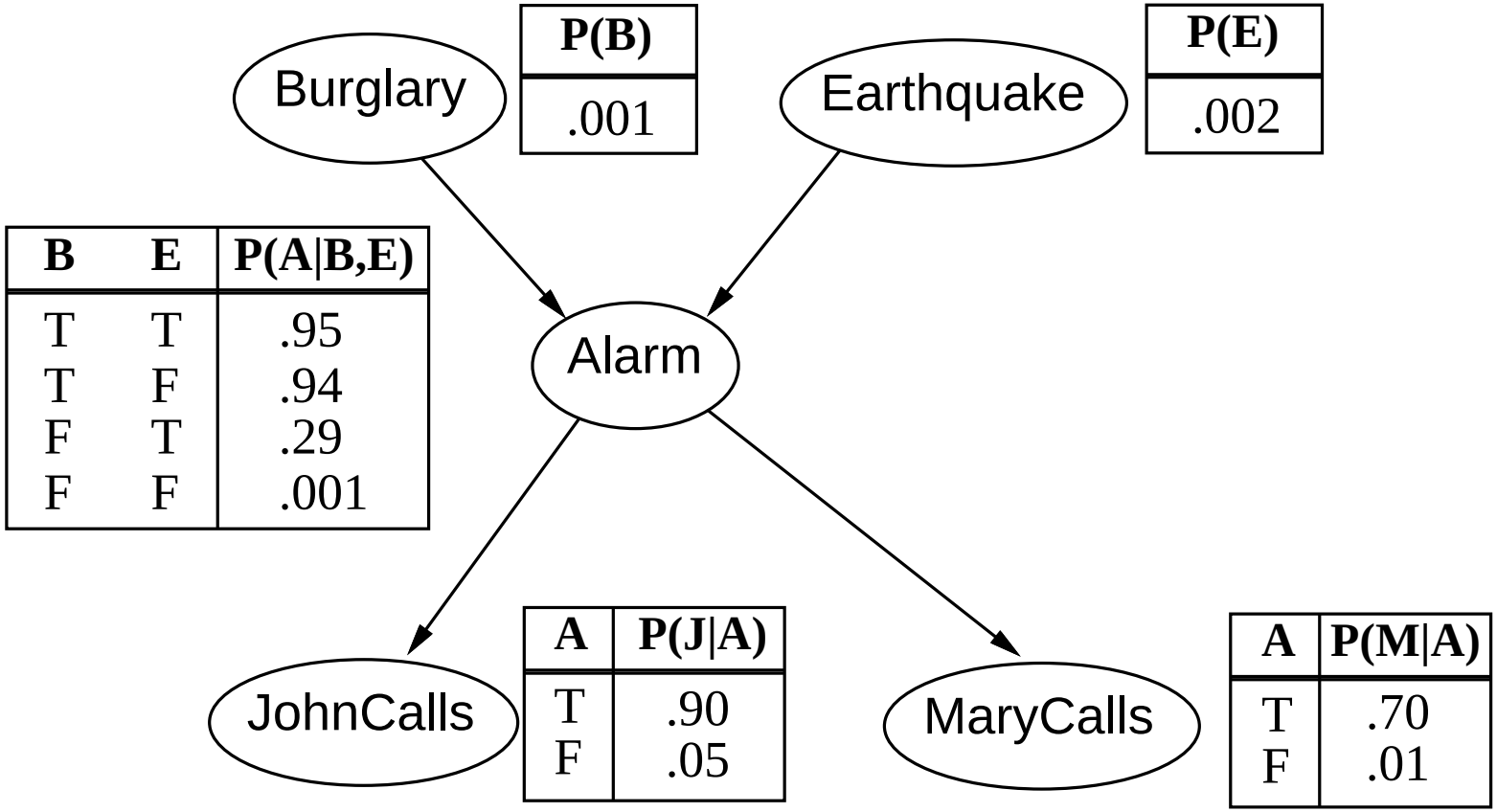
# Sieci bayessowskie: przyklad

Jestem w pracy, sąsiad John dzwoni do mnie, mówiąc mi, że mój alarm domowy się włączył, ale sąsiadka Mary nie dzwoni. Czasami alarm włącza się przy drobnych trzęsieniach ziemi. Czy to jest włamanie?

Zmienne: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
Topologia sieci odzwierciedla wiedzę "przyczynowo-skutkową":

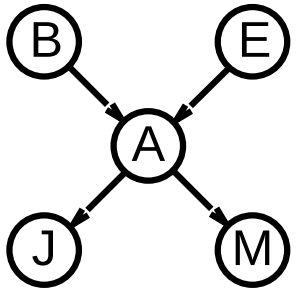
- Włamywacz może uruchomić alarm
- Trzęsienie ziemi może uruchomić alarm
- Uruchomiony alarm może spowodować, że Mary zadzwoni
- Uruchomiony alarm może spowodować, że John zadzwoni

# Sieci bayessowskie: przyklad



# Zwartosc reprezentacji sieci

TPW dla boolowskiej zmiennej $X_i$
z $k$ boolowskimi zmiennymi-rodzicami
ma $2^k$ wierszy będących kombinacjami
wartości zmiennych-rodziców




Każdy wiersz TPW wymaga
jednej wartości prawd. $p$ dla $X_i = true$
(prawdopodbieństwo dla $X_i = false$ jest $1 - p$)

Jeśli każda zmienna ma co najwyżej $k$ rodziców,
to pełna sieć wymaga $O(n \cdot 2^k)$ wartości prawdopodobieństw

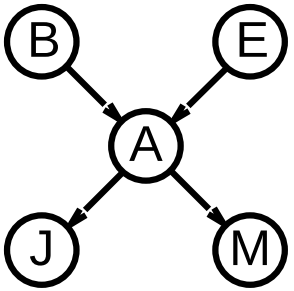
Tzn. rośnie liniowo z $n$, vs. $O(2^n)$ dla pełnego rozkładu łącznego

Dla sieci z włamaniem, $1 + 1 + 4 + 2 + 2 = 10$ wartości prawdopodbieństw
(vs. $2^5 - 1 = 31$)

# Globalna semantyka

Globalna semantyka definiuje pełny rozkład łączny jako produkt lokalnych rozkładów warunkowych:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$




np. $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$=$

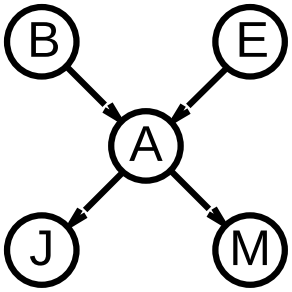
# Globalna semantyka

Globalna semantyka definiuje pełny rozkład łączny jako produkt lokalnych rozkładów warunkowych:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

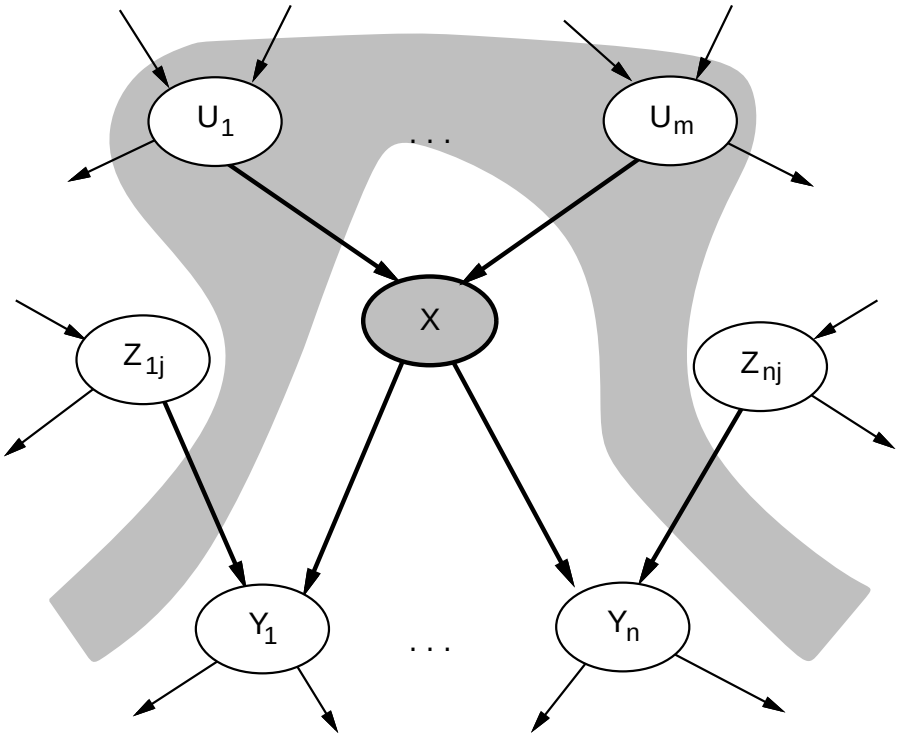
np. $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e)$$



# Lokala semantyka

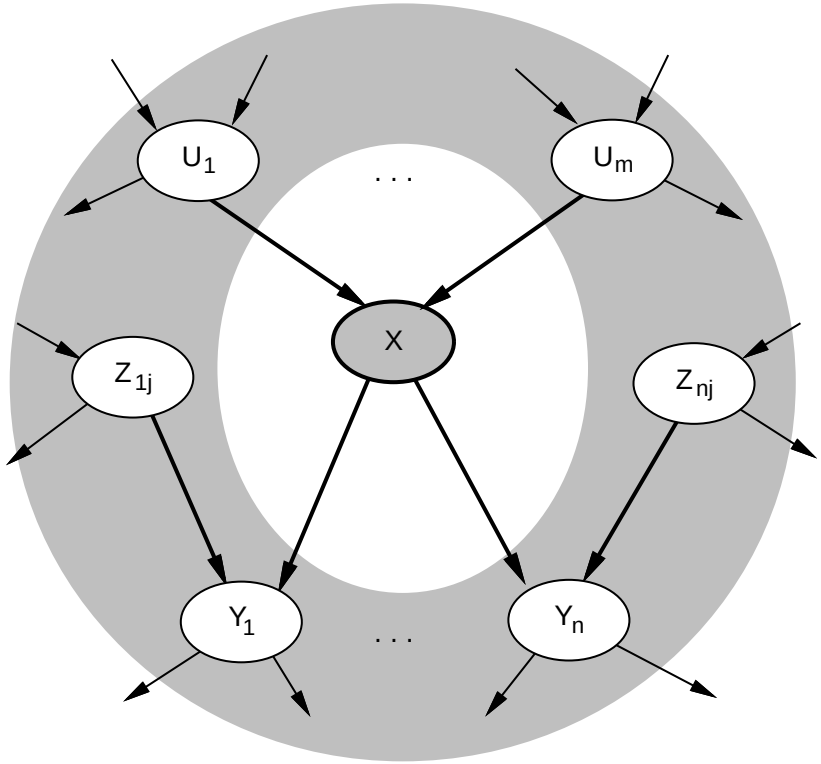
Lokalna semantyka: każdy węzeł jest warunkowo niezależny
przy danych rodzicach od pozostałych węzłów nie będących jego potomkami




Twierdzenie: Lokalna semantyka $\Leftrightarrow$ globalna semantyka

# Koc Markowa

Każdy węzeł jest warunkowo niezależny od wszystkich pozostałych przy danym jego kocu Markowa: rodzice + dzieci + inni rodzice dzieci



# Konstruowanie sieci bayessowskiej

Wymaga metody takiej, że ciąg lokalnie testowalnych zależności warunkowych nadaje znaczenie globalne

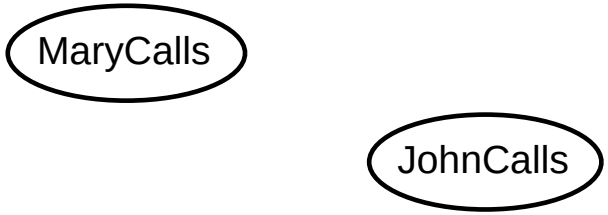
1. Wybierz uporządkowanie zmiennych los. $X_1, \ldots, X_n$
2. Dla każdego $i = 1$ do $n$
   dodaj $X_i$ do sieci
   wybierz rodziców $X_1, \ldots, X_{i-1}$ takich, że
   $$\mathbf{P}(X_i|Parents(X_i)) = \mathbf{P}(X_i|X_1, \ldots, X_{i-1})$$

Wybór rodziców gwarantuje znaczenie globalne:

$$\begin{aligned} \mathbf{P}(X_1, \ldots, X_n) &= \Pi_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1}) \quad \text{(reguła łańcuchowa)} \\ &= \Pi_{i=1}^{n} \mathbf{P}(X_i|Parents(X_i)) \quad \text{(przez konstrukcję)} \end{aligned}$$

# Konstruowanie sieci bayessowskiej: przyklad

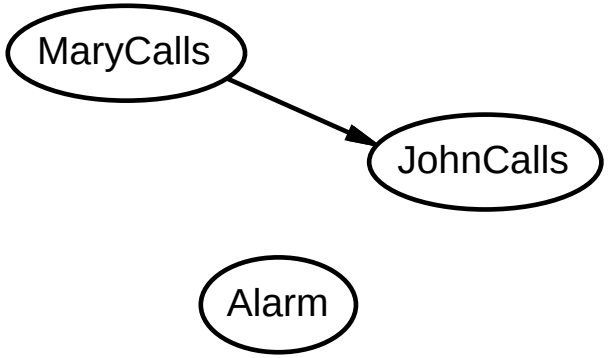
Załóżmy, że wybieramy $M$, $J$, $A$, $B$, $E$




$P(J|M) = P(J)$?

# Konstruowanie sieci bayessowskiej: przyklad

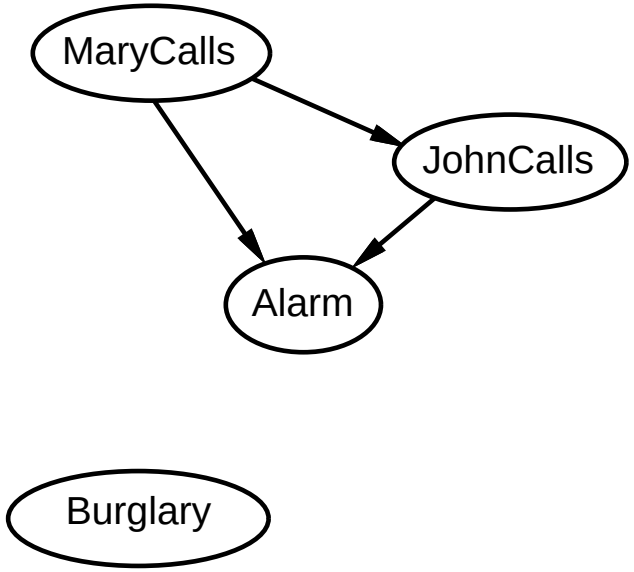
Załóżmy, że wybieramy $M$, $J$, $A$, $B$, $E$




$P(J|M) = P(J)$? Nie
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$?

# Konstruowanie sieci bayessowskiej: przyklad

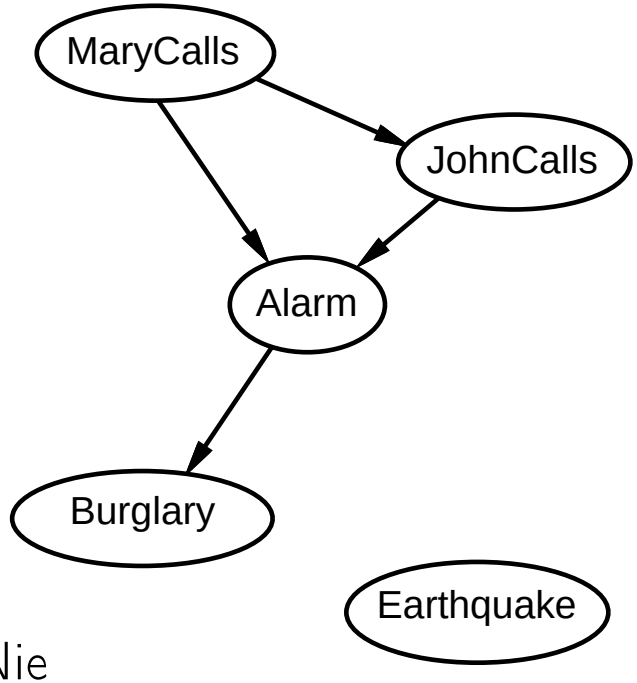
Załóżmy, że wybieramy $M$, $J$, $A$, $B$, $E$




$P(J|M) = P(J)$? Nie
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? Nie
$P(B|A, J, M) = P(B|A)$?
$P(B|A, J, M) = P(B)$?

# Konstruowanie sieci bayessowskiej: przyklad

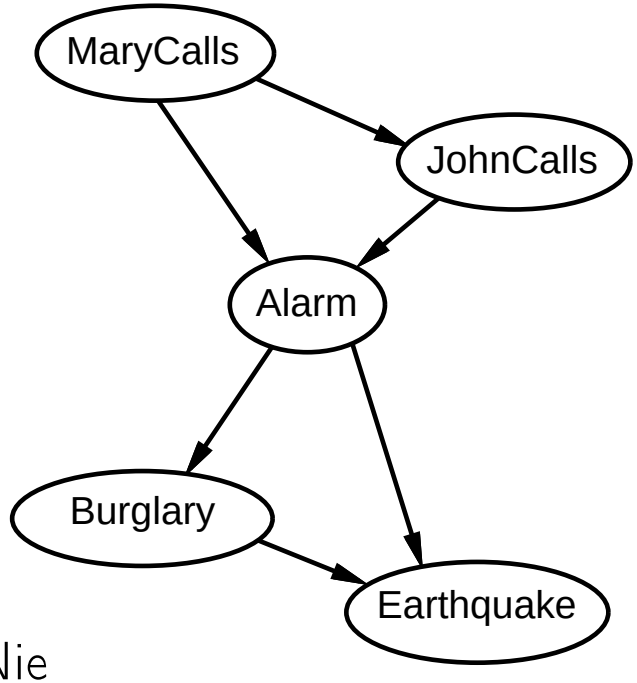
Załóżmy, że wybieramy $M$, $J$, $A$, $B$, $E$




$P(J|M) = P(J)$? Nie
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? Nie
$P(B|A, J, M) = P(B|A)$? Tak
$P(B|A, J, M) = P(B)$? Nie
$P(E|B, A, J, M) = P(E|A)$?
$P(E|B, A, J, M) = P(E|A, B)$?

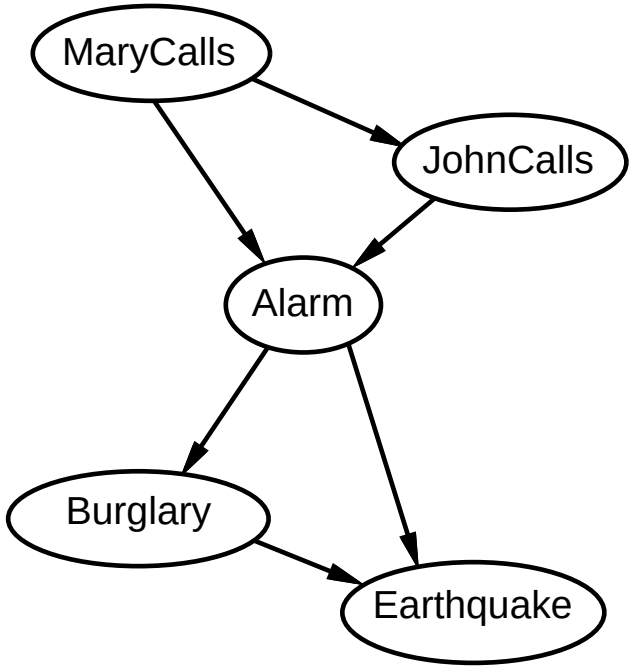
# Konstruowanie sieci bayessowskiej: przyklad

Załóżmy, że wybieramy $M$, $J$, $A$, $B$, $E$




$P(J|M) = P(J)$? Nie
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? Nie
$P(B|A, J, M) = P(B|A)$? Tak
$P(B|A, J, M) = P(B)$? Nie
$P(E|B, A, J, M) = P(E|A)$? Nie
$P(E|B, A, J, M) = P(E|A, B)$? Tak

# Konstruowanie sieci bayessowskiej: przyklad

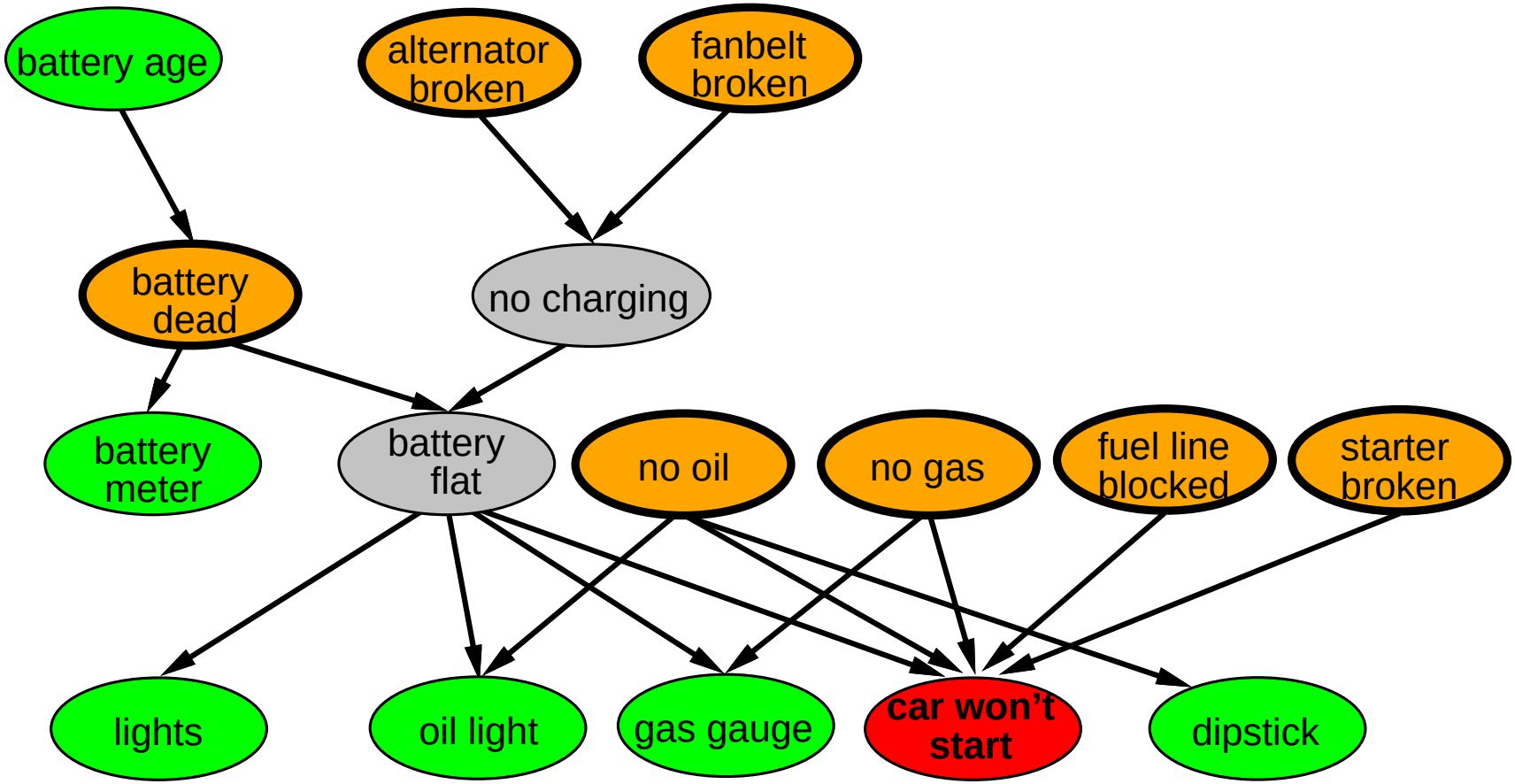



Rozpoznawanie warunkowych niezależności i oszacowanie prawdopodbieństw warunkowych jest trudne dla ludzi w nie przyczynowo-skutkowych kierunkach

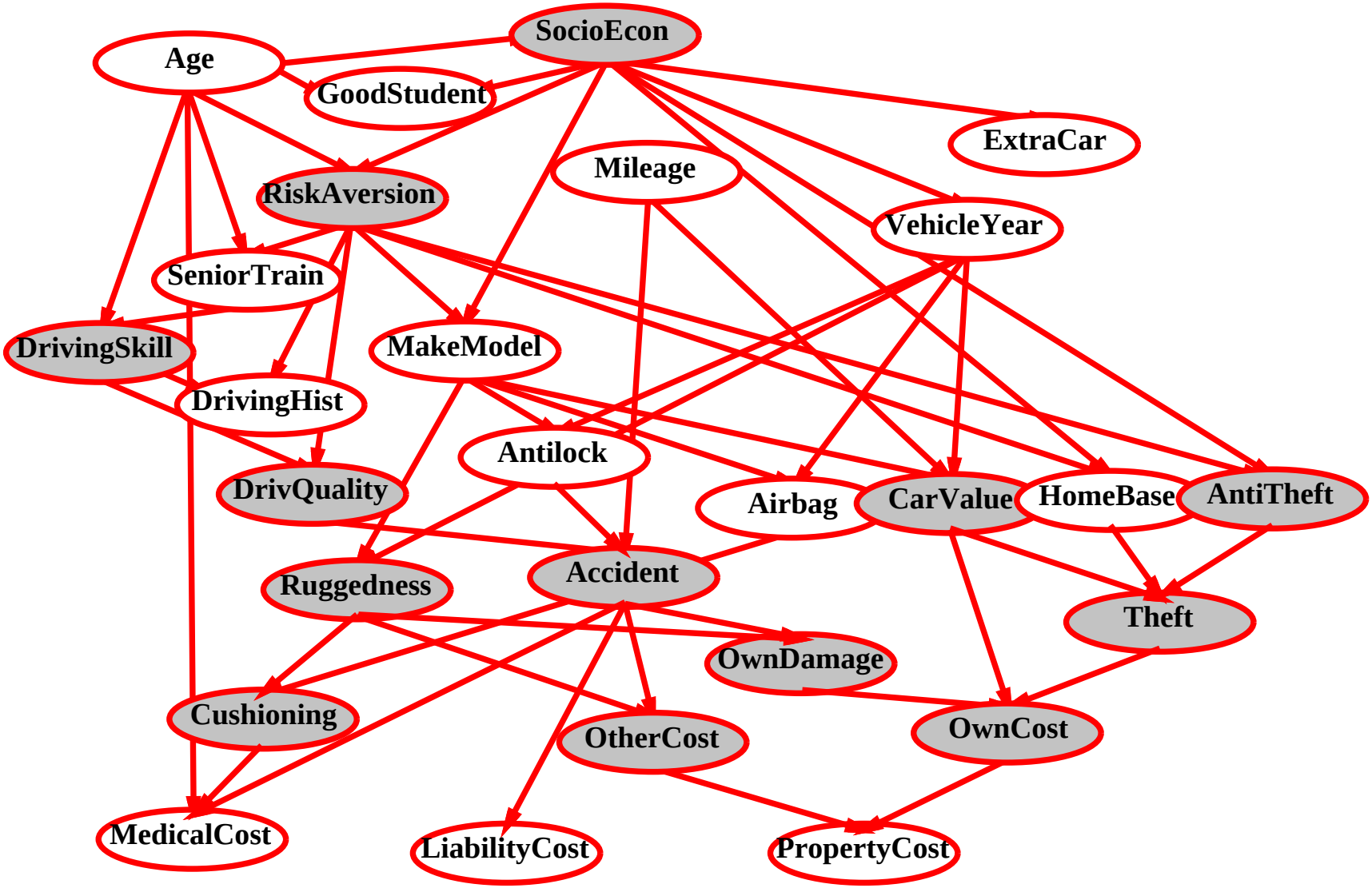
Sieć jest mniej zwarta: $1+2+4+2+4=13$ wartości prawdopodobieństw jest potrzebne

# Siec bayessowska: diagnoza samochodu

Początkowa przesłanka: samochód nie zapala
Zmienne testowalne (zielone), zmienne "zepsute, napraw to" (pomarańczowe), zmienne ukryte (szare) rozrzedzają strukturę, redukują parametry



# Siec bayessowska: ubezpieczenie samochodu



# Zwarty rozklad warunkowy

TPW rośnie wykładniczo wraz z liczbą zmiennych-rodziców
TPW staje się nieskończona dla rodzica lub syna z wartościami ciągłymi

Rozwiązanie: kanoniczne rozkłady, które są zdefiniowane w zwarty sposób

Determinstyczne węzły są najprostszym przypadkiem:
$X = f(Parents(X))$ dla pewnej funkcji $f$

Np. funkcje boolowskie
$NorthAmerican \Leftrightarrow Canadian \vee US \vee Mexican$

Np. numeryczne powiązania pomiędzy zmiennymi ciągłymi

$$\frac{\partial Level}{\partial t} = \text{inflow} + \text{precipitation} - \text{outflow} - \text{evaporation}$$

# Zwarty rozklad warunkowy

Rozkłady noisy-OR modelują wiele niezależnych przyczyn

1) Rodzice $U_1 \ldots U_k$ obejmują wszystkie możliwe przyczyny
2) Niezależne prawdopodobieństwo porażki $q_i$ dla każdej przyczyny

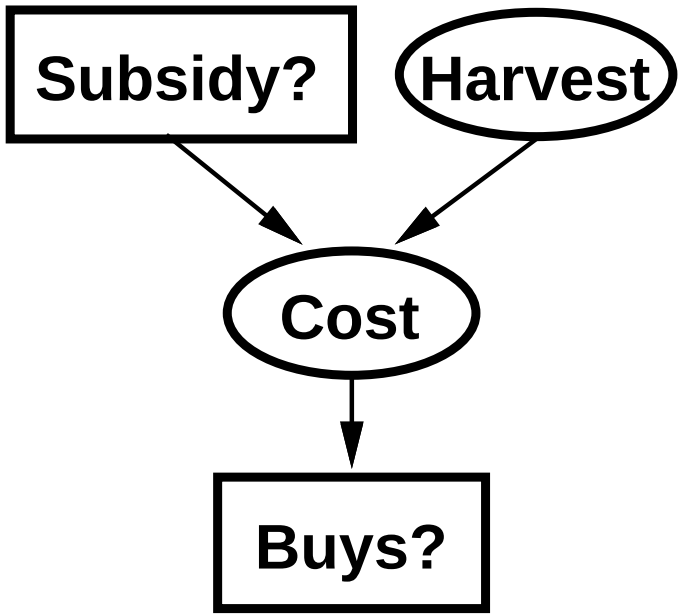
$$\Rightarrow P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \Pi_{i=1}^{j} q_i$$

| $Cold$ | $Flu$ | $Malaria$ | $P(Fever)$ | $P(\neg Fever)$ |
|---|---|---|---|---|
| F | F | F | **0.0** | 1.0 |
| F | F | T | 0.9 | **0.1** |
| F | T | F | 0.8 | **0.2** |
| F | T | T | 0.98 | $0.02 = 0.2 \times 0.1$ |
| T | F | F | 0.4 | **0.6** |
| T | F | T | 0.94 | $0.06 = 0.6 \times 0.1$ |
| T | T | F | 0.88 | $0.12 = 0.6 \times 0.2$ |
| T | T | T | 0.988 | $0.012 = 0.6 \times 0.2 \times 0.1$ |

Liczba parametrów **liniowa** od liczby rodziców

# Sieci hybrydowe (zmienne dyskretne+ciagle)

Dyskretne (*Subsidy?* i *Buys?*); ciągłe (*Harvest* i *Cost*)




Opcja 1: dyskretyzacja zm. ciągłych — możliwe duże błedy, duże TPW
Opcja 2: skończenie parametryzowalne rodziny funkcji kanonicznych

1) Zmienne ciągłe, zmienne-rodzice dyskretne+ciągłe (np. *Cost*)
2) Zmienne dyskretne, zmienne-rodzice ciągłe (np. *Buys?*)

# Zmienne-dzieci ciagle

Wymaga jednej funkcji warunkowej gęstości dla zmiennej będącej dzieckiem przy ciągłych zmiennych-rodzicach, dla każdego możliwego przypisania na zmiennych-rodzicach dyskretnych

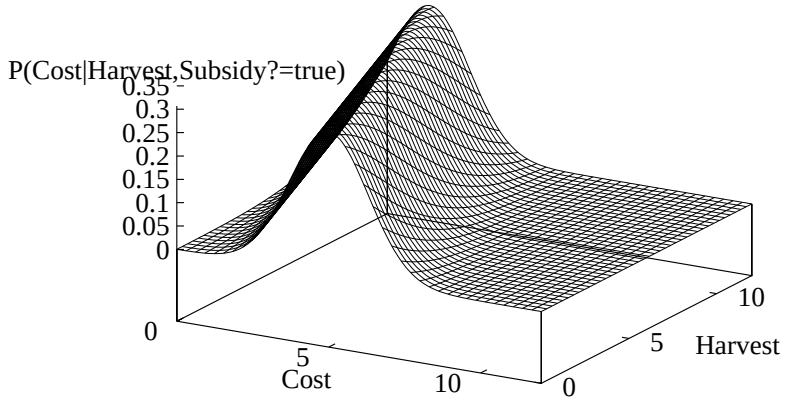
Najbardziej powszechny jest model gaussowski liniowy (LG), np.:

$$\begin{aligned} & P(Cost = c | Harvest = h, Subsidy? = true) \\ &= N(a_t h + b_t, \sigma_t)(c) \\ &= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c - (a_t h + b_t)}{\sigma_t}\right)^2\right) \end{aligned}$$

Średnia zmiennej $Cost$ zmienia się liniowo w zależności od wartości $Harvest$, wariancja jest stała

Liniowa zmienność jest nieodpowiednia dla pełnego zakresu wartości $Harvest$
ale działa dobrze, jeśli **prawdopodobny** zakres tych wartości jest wąski

# Zmienne-dzieci ciagle



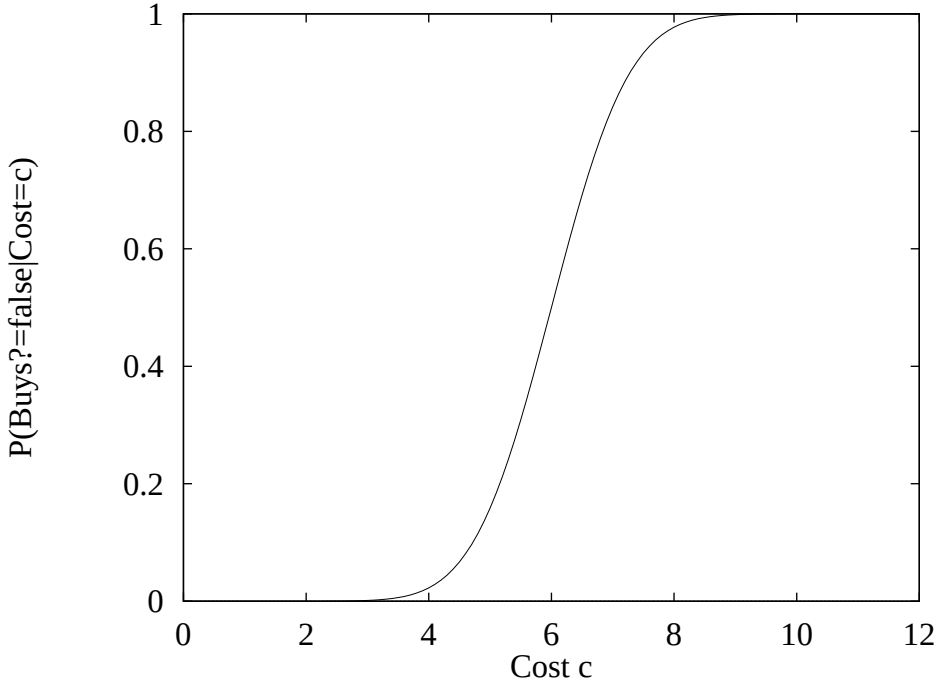

Sieć tolko ze zmiennymi ciągłymi z rozkładami LG

$\Rightarrow$ pełny rozkład gaussowski wielu zmiennych

Sieć LG zmiennych dyskretnych+ciągłych jest siecią gaussowską warunkową tzn. gaussowski rozkład wszystkich zmiennych ciągłych dla każdej kombinacji wartości zmiennych dyskretnych

# Zm. dyskretne z ciaglymi zmiennymi-rodzicami

Prawdopodob. $Buys?$ dla danego $Cost$ powinno być "miękkim" progiem:




Rozkład probitowy używa całkowania funkcji gaussowskiej:

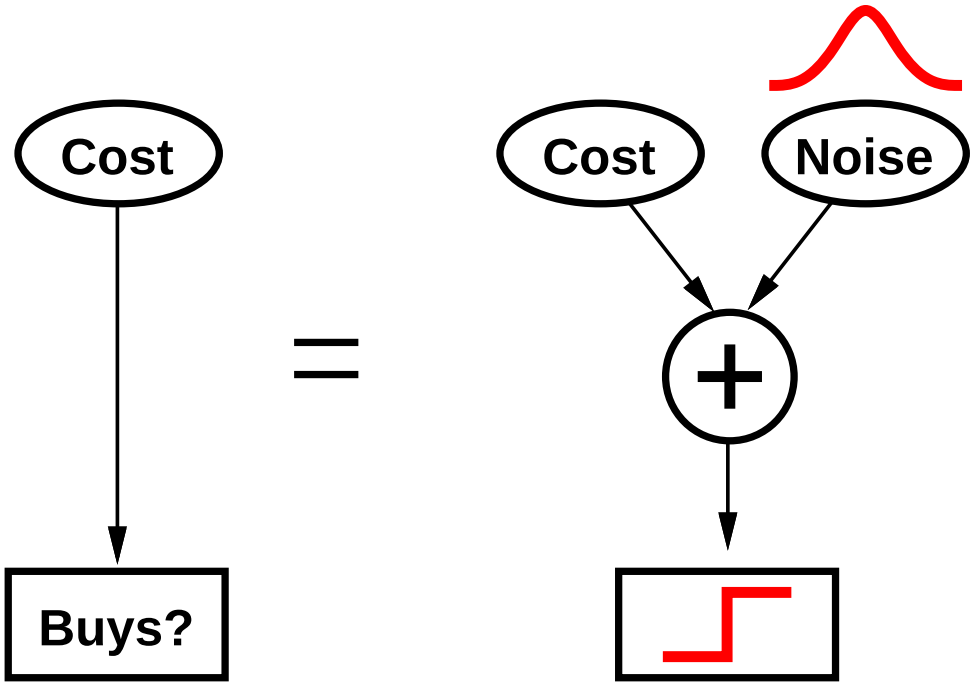
$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx$

$P(Buys? = true \mid Cost = c) = \Phi((-c + \mu)/\sigma)$

# Dlaczego rozklad probitowy?

1. Ma właściwy kształt

2. Może być traktowany jako sztywny próg, którego położenie jest zakłócone



# Zm. dyskretne z ciaglymi zmiennymi-rodzicami

Rozkład sigmoidalny (lub logitowy) używany również w siecach neuronowych:

$$P(Buys? = true \mid Cost = c) = \frac{1}{1 + exp(-2\frac{-c+\mu}{\sigma})}$$

Rozkład sigmoidalny ma kształt podobny do probitowego, ale dłuższe ogony: