Wnioskowanie Probabilistyczne

Na podstawie: AIMA, ch14

Wojciech Jaśkowski

Instytut Informatyki,
Politechnika Poznańska

3 kwietnia 2014

2014-04-03

Wnioskowanie Probabilistyczne

Wnioskowanie Probabilistyczne
Na podstawie: AIMA, ch14

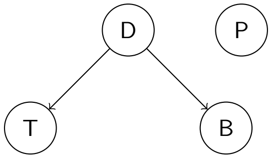
Wojciech Jaśkowski

Instytut Informatyki,
Politechnika Poznańska

3 kwietnia 2014

Na poprzednim wykładzie

	ból		*¬ból*	
	test	*¬test*	*test*	*¬test*
dziura	0.108	0.012	0.072	0.008
¬dziura	0.016	0.064	0.144	0.576

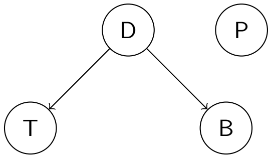


1. Łączny rozkład prawd. pozwala odpowiedzieć na każde pytanie. Ale:
   - liczba „danych" (wart. prawd) rośnie wykładniczo z liczbą zmiennych losowych
   - podawanie tych prawd. jest nienaturalne

Na poprzednim wykładzie
Łączny rozkład prawd. pozwala odpowiedzieć na każde pytanie. Ale:
liczba „danych" (wart. prawd) rośnie wykładniczo z liczbą zmiennych losowych
podawanie tych prawd. jest nienaturalne

Na poprzednim wykładzie

	ból		*¬ból*	
	test	*¬test*	*test*	*¬test*
dziura	0.108	0.012	0.072	0.008
¬dziura	0.016	0.064	0.144	0.576



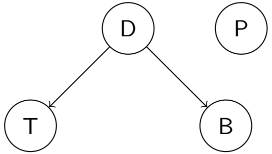
1. Łączny rozkład prawd. pozwala odpowiedzieć na każde pytanie. Ale:
   - liczba „danych" (wart. prawd) rośnie wykładniczo z liczbą zmiennych losowych
   - podawanie tych prawd. jest nienaturalne
2. **Niezależność** i **niezależność warunkowa** może znacznie zmniejszyć liczbę potrzebnych wartości:

$$\mathbf{P}(P, D) = \mathbf{P}(P)\mathbf{P}(D)$$
$$\mathbf{P}(B, T|D) = \mathbf{P}(B|D)\mathbf{P}(T|D)$$

Na poprzednim wykładzie
Łączny rozkład prawd. pozwala odpowiedzieć na każde pytanie. Ale:
liczba „danych" (wart. prawd) rośnie wykładniczo z liczbą zmiennych losowych
podawanie tych prawd. jest nienaturalne
Niezależność i niezależność warunkowa może znacznie zmniejszyć liczbę potrzebnych wartości:
P(P, D) = P(P)P(D)
P(B, T|D) = P(B|D)P(T|D)

Sieć baysowska

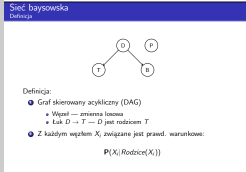
Definicja



Definicja:

1. Graf skierowany acykliczny (DAG)
   - Węzeł — zmienna losowa
   - Łuk $D \to T$ — D jest rodzicem T
2. Z każdym węzłem X_i związane jest prawd. warunkowe:

$$\mathbf{P}(X_i | Rodzice(X_i))$$



Sieć baysowska
Interpretacja

- $X_i \to X_j$ mówi, że X_i ma **bezpośredni wpływ na** X_j.
- Rozkład prawd. łącznego:

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

Sieć baysowska
Interpretacja
- $X_i \to X_j$ mówi, że X_i ma **bezpośredni wpływ na** X_j.
- Rozkład prawd. łącznego:
$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Rodzice(X_i))$$

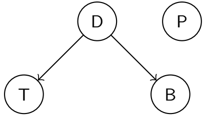
1. Ekspertowi w danej dziedzinie łatwo ten bezpośredni wpływ określić. Trudniej za to określić same prawd.

Sieć baysowska
Interpretacja

- $X_i \to X_j$ mówi, że X_i ma **bezpośredni wpływ na** X_j.
- Rozkład prawd. łącznego:

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

- Przykład:

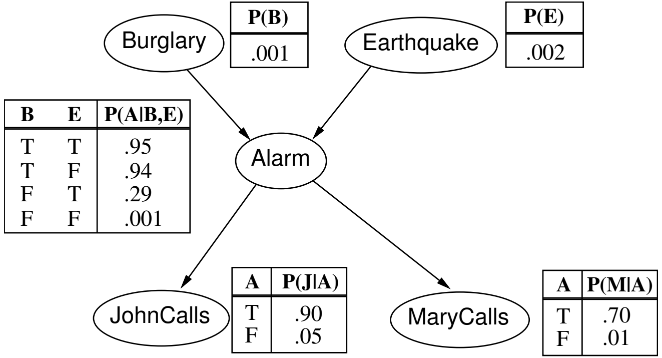


$$\begin{aligned}\mathbf{P}(D, T, B, P) &= \mathbf{P}(P)\mathbf{P}(D, T, B) \text{ (niezależność)}\\ &= \mathbf{P}(P)\mathbf{P}(T, B|D)\mathbf{P}(D) \text{ (reguła produkcji)}\\ &= \mathbf{P}(P)\mathbf{P}(B|D)\mathbf{P}(T|D)\mathbf{P}(D) \text{ (niezależność warunkow}\end{aligned}$$

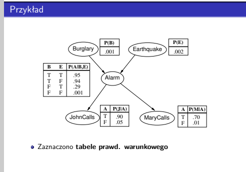
Sieć baysowska
Interpretacja
$X_i \to X_j$ mówi, że X_i ma bezpośredni wpływ na X_j.
Rozkład prawd. łącznego:
$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Rodzice(X_i))$
Przykład:
$\mathbf{P}(D, T, B, P) = \mathbf{P}(P)\mathbf{P}(D, T, B)$ (niezależność)
$= \mathbf{P}(P)\mathbf{P}(T, B|D)\mathbf{P}(D)$ (reguła produkcji)
$= \mathbf{P}(P)\mathbf{P}(B|D)\mathbf{P}(T|D)\mathbf{P}(D)$ (niezależność warunkow

1. Ekspertowi w danej dziedzinie łatwo ten bezpośredni wpływ określić. Trudniej za to określić same prawd.

Przykład

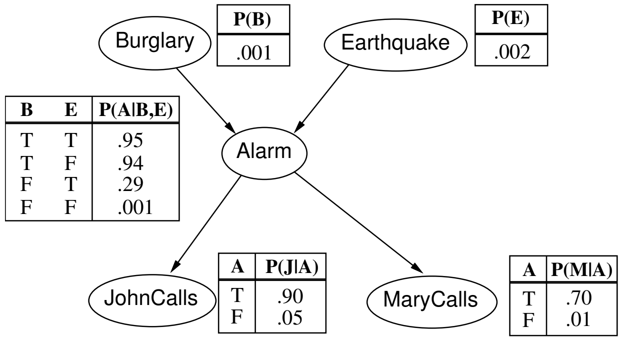


- Zaznaczono **tabele prawd. warunkowego**



1. Widać naszą ignorancję i lenistwo: Nie ma prawd., ze Marry słucha głośnej muzyki. To jest już ukryte w węźle Marry.
2. Nie uwzględnia również wielu innych czynników

Przykład

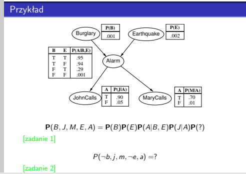


$$\mathbf{P}(B, J, M, E, A) = \mathbf{P}(B)\mathbf{P}(E)\mathbf{P}(A|B, E)\mathbf{P}(J|A)\mathbf{P}(?)$$

[zadanie 1]

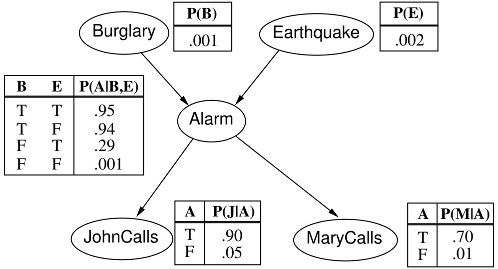
$$P(\neg b, j, m, \neg e, a) = ?$$

[zadanie 2]



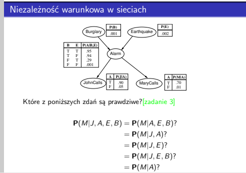
1. Odp. $\mathbf{P}(M|A)$
2. Odp. 0.000628
3. (Zauważmy, że J i M nie mają strzałki to i tak są zależne (marginalnie), tj. $J \not\perp M$, natomiast są niezależne warunkowo $J \perp M|A$).

Niezależność warunkowa w sieciach



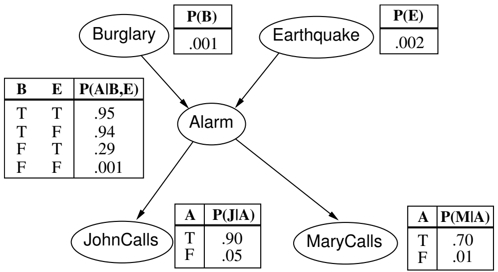
Które z poniższych zdań są prawdziwe?[zadanie 3]

$$\begin{aligned}
\mathbf{P}(M|J,A,E,B) &= \mathbf{P}(M|A,E,B)?\\
&= \mathbf{P}(M|J,A)?\\
&= \mathbf{P}(M|J,E)?\\
&= \mathbf{P}(M|J,E,B)?\\
&= \mathbf{P}(M|A)?
\end{aligned}$$

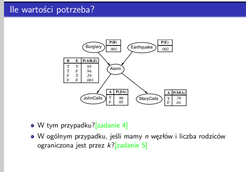


1. $\mathbf{P}(M|J,A,E,B) = \mathbf{P}(M|A) = \mathbf{P}(M|A,E,B) = \mathbf{P}(M|J,A)$
2. I właśnie $\mathbf{P}(M|J,A,E,B) = \mathbf{P}(M|A)$ jest cechą sieci — wystarczy znać rodzica i reszta się wyjaśnia

Ile wartości potrzeba?



- W tym przypadku?[zadanie 4]
- W ogólnym przypadku, jeśli mamy n węzłów i liczba rodziców ograniczona jest przez k?[zadanie 5]



1. 10
2. $\leq n2^k$

Niezależność zwykła vs. niezależność warunkowa

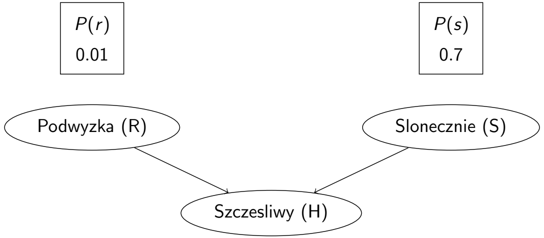
- $T_1 \perp T_2 | C \implies T_1 \perp T_2$?[zadanie 6]
- $T_1 \perp T_2 \implies T_1 \perp T_2 | C$?[zadanie 7]

Niezależność zwykła vs. niezależność warunkowa

- $T_1 \perp T_2 | C \implies T_1 \perp T_2$?[zadanie 6]
- $T_1 \perp T_2 \implies T_1 \perp T_2 | C$?[zadanie 7]

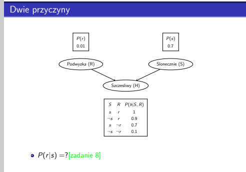
1. Nie. Przykład: rak i dwa testy. Jest niezależność warunkowa, ale nie ma niezależności absolutnej, bo $P(t_1|t_2) > P(t_1)$ (czyli, pozytywny wynik testu T_2 wpływa na prawd., że otrzymamy pozytywny wynik testu T_1.
2. Też nie. I zaraz to wyjaśnimy

Dwie przyczyny



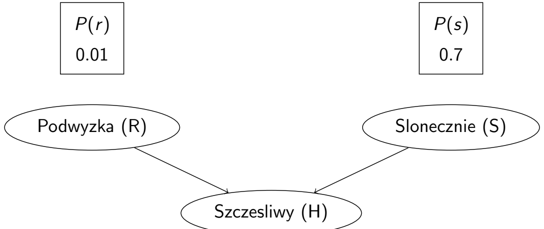
S	R	$P(h \mid S, R)$
s	r	1
$\neg s$	r	0.9
s	$\neg r$	0.7
$\neg s$	$\neg r$	0.1

- $P(r|s) = ?$[zadania 8]



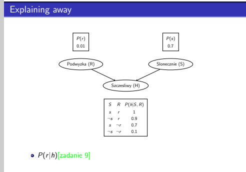
1. 0.01, bo R i S są niezależne (bezwarunkowo!)

Explaining away



S	R	$P(h \mid S, R)$
s	r	1
$\neg s$	r	0.9
s	$\neg r$	0.7
$\neg s$	$\neg r$	0.1

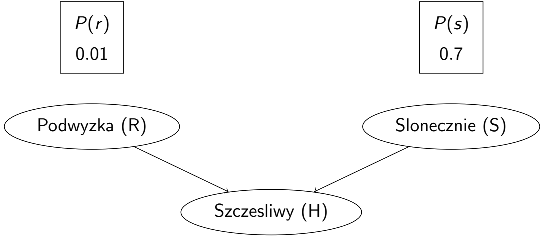
- $P(r \mid h)$[zadanie 9]



1.

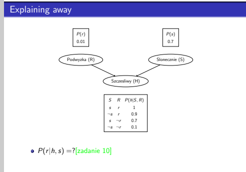
$$\mathbf{P}(R \mid h) = \alpha \mathbf{P}(R, h) = \alpha \sum_s \mathbf{P}(R, h, s) \text{ (prawd. całkowite)}$$
$$= \alpha \sum_s \mathbf{P}(h \mid s, R) \mathbf{P}(R, s) \text{ (reguła produkcji)}$$
$$= \alpha \sum_s \mathbf{P}(h \mid s, R) \mathbf{P}(R) \mathbf{P}(s) \text{ (niezależność)}$$
$$= \alpha \mathbf{P}(R) \sum_s \mathbf{P}(h \mid s, R) \mathbf{P}(s) \text{ (wyciągnięcie przed nawias)}$$
$$= \langle 0.01 \times (1 \times 0.7 + 0.9 \times 0.3), 0.99 \times (0.7 \times 0.7 + 0.1 \times 0.3) \rangle$$
$$= \langle 0.0185, 0.9815 \rangle$$

Explaining away



S	R	$P(h \mid S, R)$
s	r	1
$\neg s$	r	0.9
s	$\neg r$	0.7
$\neg s$	$\neg r$	0.1

- $P(r \mid h, s) = ?$[zadanie 10]

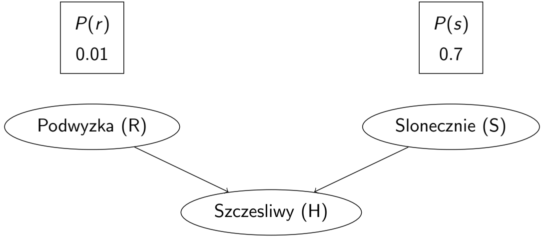


1.

$$\begin{aligned}\mathbf{P}(R \mid h, s) &= \alpha \mathbf{P}(h \mid R, s)\mathbf{P}(R \mid s) \text{ (uwaga: ugólniony Bayes)}\\ &= \alpha \mathbf{P}(h \mid R, s)\mathbf{P}(R) \text{ (niezależność } R \text{i } S\text{)}\\ &= \langle 0.0142, 0.9858\rangle\end{aligned}$$

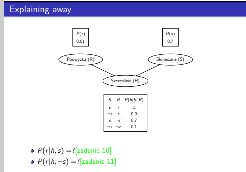
2. Analogicznie jak $P(r \mid h, s)$. Wychodzi: 0.0833.

Explaining away



S	R	$P(h\|S,R)$
s	r	1
$\neg s$	r	0.9
s	$\neg r$	0.7
$\neg s$	$\neg r$	0.1

- $P(r|h,s) = ?$[zadanie 10]
- $P(r|h,\neg s) = ?$[zadanie 11]

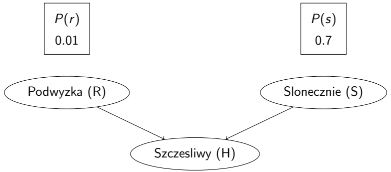


1.

$$\begin{aligned}\mathbf{P}(R|h,s) &= \alpha\mathbf{P}(h|R,s)\mathbf{P}(R|s) \text{ (uwaga: ugólniony Bayes)}\\ &= \alpha\mathbf{P}(h|R,s)\mathbf{P}(R) \text{ (niezależność } R \text{i } S)\\ &= \langle 0.0142, 0.9858\rangle\end{aligned}$$

2. Analogicznie jak $P(r|h,s)$. Wychodzi: 0.0833.

Explaining away — podsumowanie



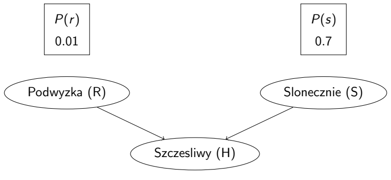
S	R	$P(h \mid S, R)$
s	r	1
$\neg s$	r	0.9
s	$\neg r$	0.7
$\neg s$	$\neg r$	0.1

$$P(r|s) = P(r) = 0.01$$
$$P(r|h) = 0.0185$$
$$P(r|h, s) = 0.0142$$
$$P(r|h, \neg s) = 0.0833$$

- Jeśli wiemy, że jesteśmy radośni, pogoda może wyjaśnić przyczynę radości. Jeśli jednak okazuje się, że jest słonecznie, to prawd., że dostałem podwyżkę znacznie maleje.
- Wniosek: R i S są **niezależne**, ale H powoduje, że **stają się zależne** (pod warunkiem, że H)!

Explaining away — podsumowanie
$P(r|s) = P(r) = 0.01$
$P(r|h) = 0.0185$
$P(r|h, s) = 0.0142$
$P(r|h, \neg s) = 0.0833$
Jeśli wiemy, że jesteśmy radośni, pogoda może wyjaśnić przyczynę radości. Jeśli jednak okazuje się, że jest słonecznie, to prawd., że dostałem podwyżkę znacznie maleje.
Wniosek: R i S są niezależne, ale H powoduje, że stają się zależne (pod warunkiem, że H)!

Explaining away — podsumowanie



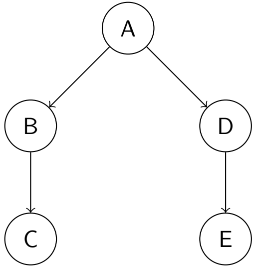
S	R	$P(h \mid S, R)$
s	r	1
$\neg s$	r	0.9
s	$\neg r$	0.7
$\neg s$	$\neg r$	0.1

$$P(r|s) = P(r) = 0.01$$
$$P(r|h) = 0.0185$$
$$P(r|h, s) = 0.0142$$
$$P(r|h, \neg s) = 0.0833$$

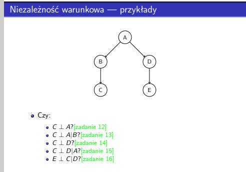
- Jeśli nie znamy H, to $P(r|s) = P(r) \rightarrow R$ **i** S **są niezależne,**
- Jeśli znamy H, to R **i** S **stają się zależne** $(P(r|h, s) \neq P(r|h))$
- Wniosek: niezależność nie implikuje niezależności warunkowej.

Explaining away — podsumowanie
$P(r|s) = P(r) = 0.01$
$P(r|h) = 0.0185$
$P(r|h, s) = 0.0142$
$P(r|h, \neg s) = 0.0833$
Jeśli nie znamy H, to $P(r|s) = P(r) \rightarrow R$ i S są niezależne,
Jeśli znamy H, to R i S stają się zależne $(P(r|h, s) \neq P(r|h))$
Wniosek: niezależność nie implikuje niezależności warunkowej.

Niezależność warunkowa — przykłady

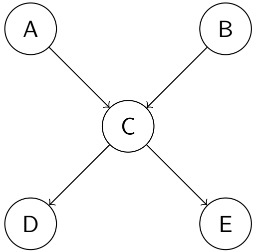


- Czy:
  - $C \perp A$?[zadanie 12]
  - $C \perp A|B$?[zadanie 13]
  - $C \perp D$?[zadanie 14]
  - $C \perp D|A$?[zadanie 15]
  - $E \perp C|D$?[zadanie 16]

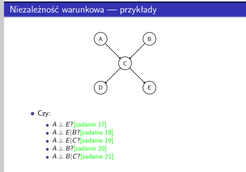


1. Nie
2. Tak
3. Nie
4. Tak
5. Tak

Niezależność warunkowa — przykłady



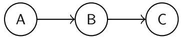
- Czy:
  - $A \perp E$?[zadanie 17]
  - $A \perp E|B$?[zadanie 18]
  - $A \perp E|C$?[zadanie 19]
  - $A \perp B$?[zadanie 20]
  - $A \perp B|C$?[zadanie 21]



1. Nie
2. Nie
3. Tak
4. Tak
5. Nie (explaining away!)

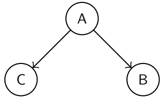
Niezależność warunkowa (zasady ogólne uproszczone)

Seryjne



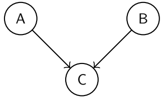
A i C są zależne, ale $A \perp C | B$

Rozchodzące się

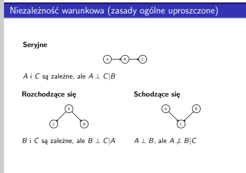


B i C są zależne, ale $B \perp C | A$

Schodzące się



$A \perp B$, ale $A \not\perp B | C$



1. podobnie w przypadku ogólnym: $A, B \rightarrow X_1 \cdots \rightarrow \ldots X_n \rightarrow C$

D-separacja (zasady ogólne)

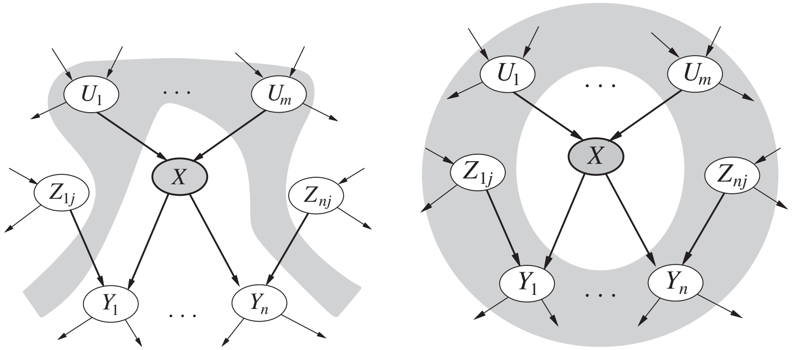
$\mathbf{X}$, $\mathbf{Y}$ i $\mathbf{Z}$ są zbiorami zmiennych losowych.

1. $\mathbf{X} \perp \mathbf{Y} | \mathbf{Z} \iff \mathbf{X}$ i $\mathbf{Y}$ są d-odseparowane.
2. $\mathbf{X}$, $\mathbf{Y}$ są **d-odseparowane** jeśli każda ścieżka prowadząca od węzła $X \in \mathbf{X}$ do węzła $Y \in \mathbf{Y}$ jest „**zablokowana**", czyli jeśli istnieje na niej jakiś węzeł $V \notin \mathbf{X} \cup \mathbf{Y}$, taki, że:
   1. Połączenie przez V jest typu **rozchodzącego się** lub **seryjnego,** oraz $V \in \mathbf{Z}$
   2. Połączenie przez V jest typu **schodzącego się** oraz ani $V \notin \mathbf{Z}$ ani żaden potomek węzła V w grafie $\notin \mathbf{Z}$

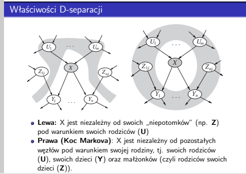
Applet do testowania d-separowalności:
`http://www.phil.cmu.edu/~wimberly/dsep/dSep.html`

D-separacja (zasady ogólne)
X, Y i Z są zbiorami zmiennych losowych.
X ⊥ Y|Z ⟺ X i Y są d-odseparowane.
X, Y są d-odseparowane jeśli każda ścieżka prowadząca od węzła X ∈ X do węzła Y ∈ Y jest „zablokowana", czyli jeśli istnieje na niej jakiś węzeł V ∉ X ∪ Y, taki, że:
Połączenie przez V jest typu rozchodzącego się lub seryjnego, oraz V ∈ Z
Połączenie przez V jest typu schodzącego się oraz ani V ∉ Z ani żaden potomek węzła V w grafie ∉ Z
Applet do testowania d-separowalności:
http://www.phil.cmu.edu/~wimberly/dsep/dSep.html

Właściwości D-separacji

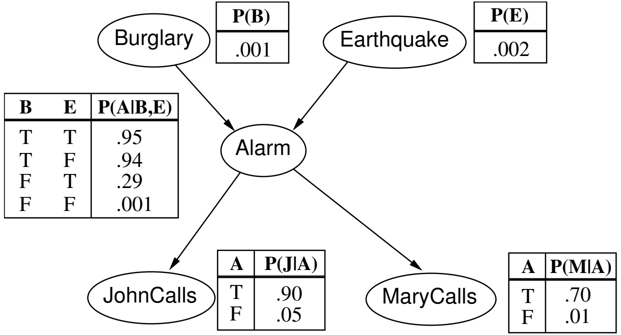


- **Lewa:** X jest niezależny od swoich „niepotomków" (np. **Z**) pod warunkiem swoich rodziców (**U**)
- **Prawa (Koc Markova)**: X jest niezależny od pozostałych węzłów pod warunkiem swojej rodziny, tj. swoich rodziców (**U**), swoich dzieci (**Y**) oraz małżonków (czyli rodziców swoich dzieci (**Z**)).



1. Lewa: Nie jest to oczywiste, ale spełnia warunki d-separacji. Należy zwrócić uwagę fakt iż, w tym przypadku także X jest niezależne (bezwarunkowo) od **Z** (o ile oczywiści nie ma krawędzi $U \to Z$, ale to zupełnie inna historia).

Wnioskowanie



P(B)
.001

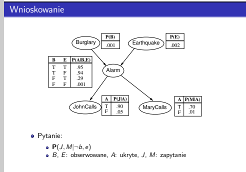
P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

- Pytanie:
  - $\mathbf{P}(J, M|\neg b, e)$
  - B, E: obserwowane, A: ukryte, J, M: zapytanie



Wnioskowanie

W ogólności pytania, które możemy stawiać mają postać:

1. **Rozkład łączny**

$$\mathbf{P}(Q_1, Q_2, \ldots | E_1 = e_1, E_2 = e_2)$$

   - wynik: łączny rozkład a posteriori (znając wartości **e**)

2. **Estymacja największej wiarygodności** (ang. *maximum likelihood estimation*):

$$argmax_{\mathbf{q}} P(Q_1 = q_1, Q_2 = q_2, \ldots | E_1 = e_1, E_2 = e_2)$$

   - wynik: wektor wartości **q** (wartości najbardziej prawdopodobne).

Wnioskowanie

W ogólności pytania, które możemy stawiać mają postać:

1. **Rozkład łączny**

$$\mathbf{P}(Q_1, Q_2, \ldots | E_1 = e_1, E_2 = e_2)$$

   - wynik: łączny rozkład a posteriori (znając wartości **e**)

2. **Estymacja największej wiarygodności** (ang. *maximum likelihood estimation*):

$$argmax_{\mathbf{q}} P(Q_1 = q_1, Q_2 = q_2, \ldots | E_1 = e_1, E_2 = e_2)$$

   - wynik: wektor wartości **q** (wartości najbardziej prawdopodobne).

Enumeracja

- Obliczamy wprost, np.

$$P(b|j,m) = P(b,j,m)/P(j,m)$$

Enumeracja
- Obliczamy wprost, np.
$P(b|j,m) = P(b,j,m)/P(j,m)$

1. Uwaga: tutaj b, j i m są stałymi, ale e i a zmiennymi (wynika to z kontekstu)

Enumeracja

- Obliczamy wprost, np.

$$P(b|j,m) = P(b,j,m)/P(j,m)$$

- obliczmy np. $P(b,j,m)$:

$$\begin{aligned} P(b,j,m) &= \sum_{e\in E}\sum_{a\in A} P(b,j,m,e,a) \text{ (prawd. całkowite)} \\ &= \sum_{e\in E}\sum_{a\in A} P(b)P(e)P(a|b,e)P(j|a)P(m|a) \end{aligned}$$

Enumeracja
- Obliczamy wprost, np.
$P(b|j,m) = P(b,j,m)/P(j,m)$
- obliczmy np. $P(b,j,m)$:
$P(b,j,m) = \sum_{e\in E}\sum_{a\in A} P(b,j,m,e,a)$ (prawd. całkowite)
$= \sum_{e\in E}\sum_{a\in A} P(b)P(e)P(a|b,e)P(j|a)P(m|a)$

1. Uwaga: tutaj b, j i m są stałymi, ale e i a zmiennymi (wynika to z kontekstu)

Enumeracja

Koszt obliczeniowy

$$P(b, j, m) = \sum_{e \in E} \sum_{a \in A} P(b)P(e)P(a|b, e)P(j|a)P(m|a)$$

- Zmienne są binarne, więc mamy $2 \times 2 = 4$ „wierszy" do policzenia
  $\Rightarrow$ w ogólności: $O(2^n)$ w najgorszym przypadku przy n zmiennych binarnych.
- Kiedy występuje najgorszy przypadek? [zadanie 22]

Enumeracja
Koszt obliczeniowy

$$P(b, j, m) = \sum_{e \in E} \sum_{a \in A} P(b)P(e)P(a|b, e)P(j|a)P(m|a)$$

- Zmienne są binarne, więc mamy $2 \times 2 = 4$ „wierszy" do policzenia
  $\Rightarrow$ w ogólności: $O(2^n)$ w najgorszym przypadku przy n zmiennych binarnych.
- Kiedy występuje najgorszy przypadek? [zadanie 22]

1. To zależy od liczby zmiennych ukrytych, więc najgorszy przypadek jest wtedy, gdy szukamy prawd. marginalnego, np. $P(m)$

Przyspieszanie enumeracji

wyciąganie przed nawias

- Niektóre wartości można wyciągnąć przed nawias:

$$
\begin{aligned}
P(b, j, m) &= \sum_{e} \sum_{a} P(b)P(e)P(a|b, e)P(j|a)P(m|a) \\
&= P(b) \sum_{e} P(e) \sum_{a} P(a|b, e)P(j|a)P(m|a)
\end{aligned}
$$

Przyspieszanie enumeracji
wyciąganie przed nawias
- Niektóre wartości można wyciągnąć przed nawias:
$$P(b, j, m) = \sum_{e} \sum_{a} P(b)P(e)P(a|b, e)P(j|a)P(m|a) = P(b) \sum_{e} P(e) \sum_{a} P(a|b, e)P(j|a)P(m|a)$$

Przyspieszanie enumeracji

wyciąganie przed nawias

- Niektóre wartości można wyciągnąć przed nawias:

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

- Niewielkie przyspieszenie $\rightarrow$ nadal zbyt wiele „wierszy"

Przyspieszanie enumeracji
wyciąganie przed nawias

- Niektóre wartości można wyciągnąć przed nawias:

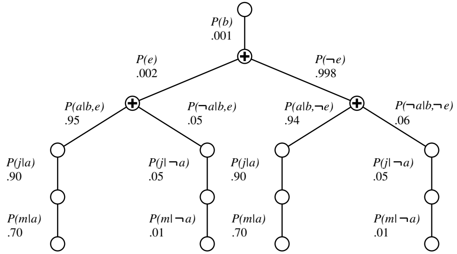
$$P(b,j,m) = \sum_e \sum_a P(b)P(e)P(a|b,e)P(j|a)P(m|a)$$
$$= P(b)\sum_e P(e)\sum_a P(a|b,e)P(j|a)P(m|a)$$

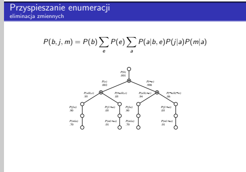
- Niewielkie przyspieszenie $\rightarrow$ nadal zbyt wiele „wierszy"

Przyspieszanie enumeracji

eliminacja zmiennych

$$P(b, j, m) = P(b) \sum_{e} P(e) \sum_{a} P(a|b, e) P(j|a) P(m|a)$$

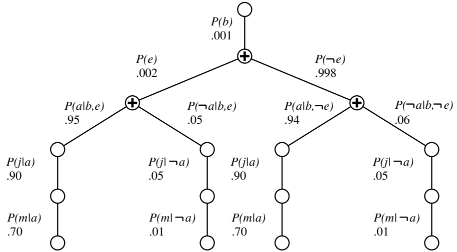




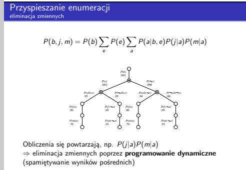
1. Niestety, to nie zmienia liczby wierszy

Przyspieszanie enumeracji
eliminacja zmiennych

$$P(b, j, m) = P(b)\sum_{e} P(e)\sum_{a} P(a|b, e)P(j|a)P(m|a)$$



Obliczenia się powtarzają, np. $P(j|a)P(m|a)$
$\Rightarrow$ eliminacja zmiennych poprzez **programowanie dynamiczne** (spamiętywanie wyników pośrednich)



1. Niestety, to nie zmienia liczby wierszy

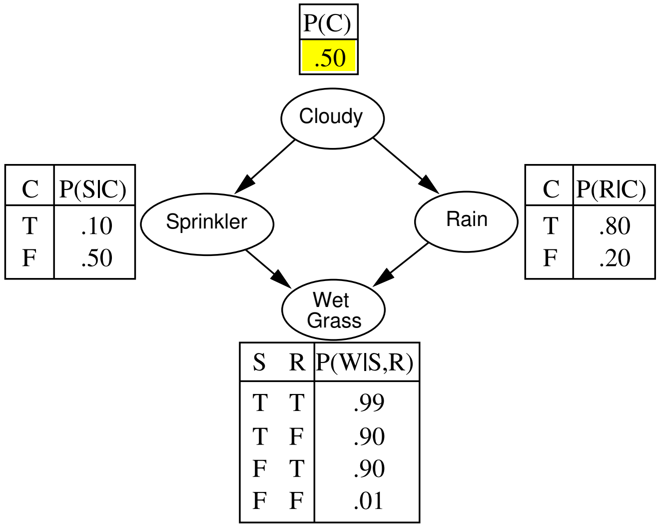
Próbkowanie bezpośrednie (ang. *direct sampling*)

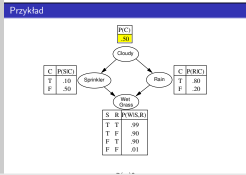
- Estymuje łączny rozkład prawdopodobieństwa
- Algorytm:
  - Wylosuj N próbek zgodnie ze znanymi prawdopodobieństwami.
  - Liczbę pasujących próbek podziel przez N

Próbkowanie bezpośrednie (ang. *direct sampling*)

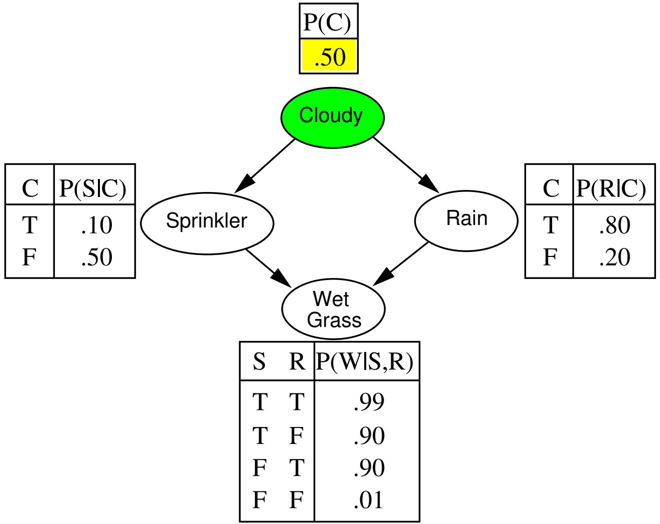
- Estymuje łączny rozkład prawdopodobieństwa
- Algorytm:
  - Wylosuj N próbek zgodnie ze znanymi prawdopodobieństwami.
  - Liczbę pasujących próbek podziel przez N

Przykład

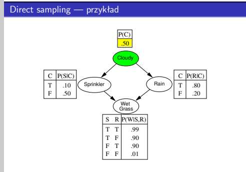

P(C)
.50
Cloudy
C P(S|C)
T .10
F .50
Sprinkler
Rain
C P(R|C)
T .80
F .20
Wet Grass
S R P(W|S,R)
T T .99
T F .90
F T .90
F F .01



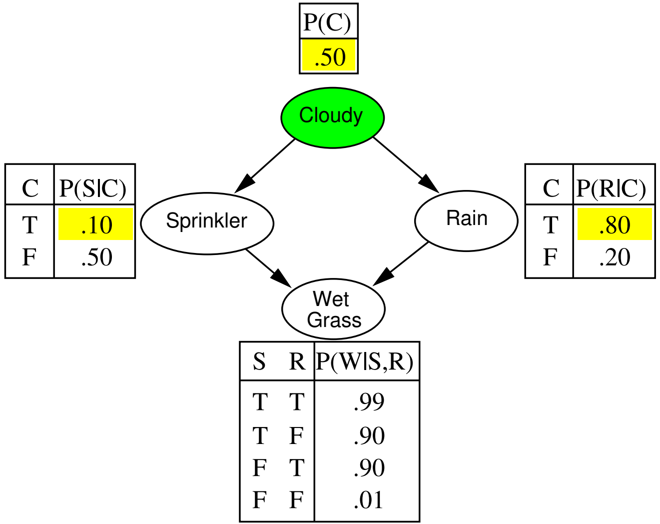
Direct sampling — przykład



2014-04-03



Direct sampling — przykład



P(C)
.50

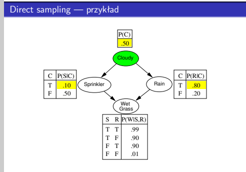
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

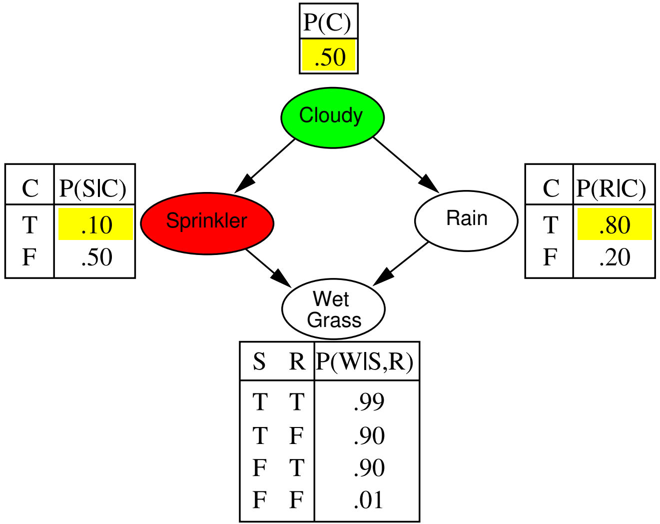
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

2014-04-03

Wnioskowanie Probabilistyczne
└Metody wnioskowania
└Direct sampling — przykład



Direct sampling — przykład

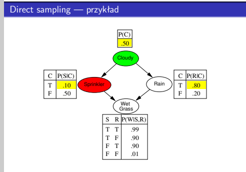


P(C)
.50

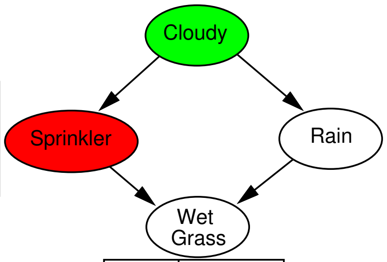
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01


Direct sampling — przykład

Direct sampling — przykład

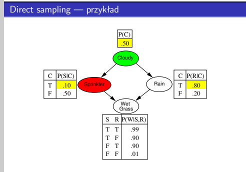


P(C)
.50

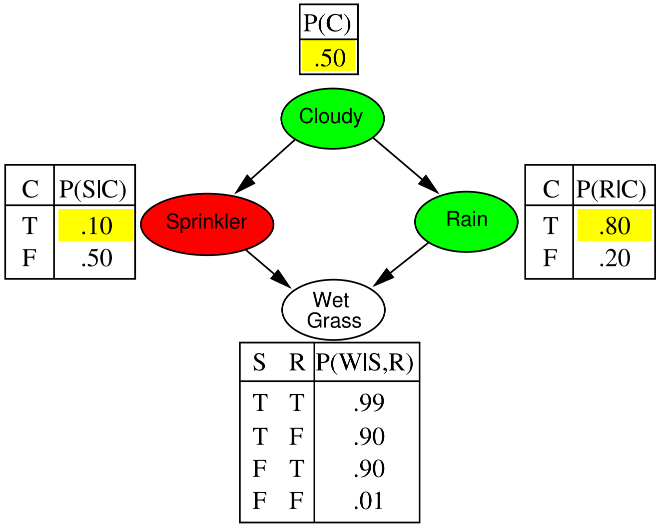
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01


Direct sampling — przykład

Direct sampling — przykład



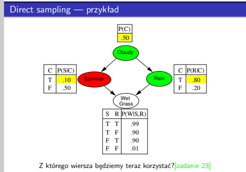
P(C)
.50

C	P(S|C)
T	.10
F	.50

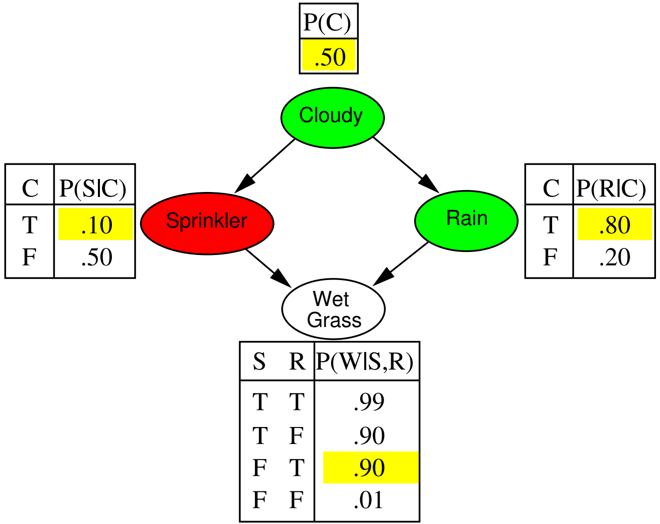
C	P(R|C)
T	.80
F	.20

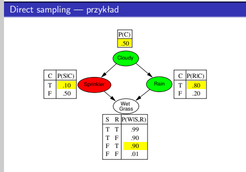
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Z którego wiersza będziemy teraz korzystać?[zadanie 23]

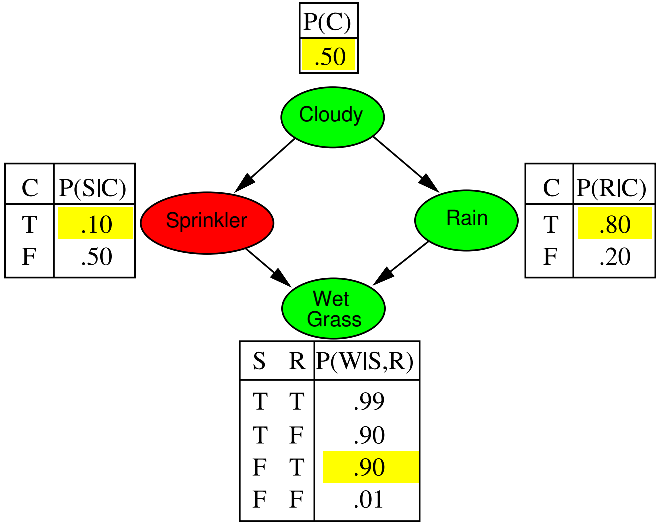


Direct sampling — przykład





Direct sampling — przykład



P(C)
.50

C	P(S|C)
T	.10
F	.50

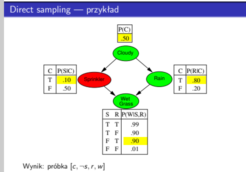
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Wynik: próbka $[c, \neg s, r, w]$

2014-04-03

Wnioskowanie Probabilistyczne
└Metody wnioskowania
└Direct sampling — przykład



Direct sampling — przykład

Wyniki kolejnych próbkowań:

1. $[c, \neg s, r, w]$
2. $[\neg c, s, \neg r, w]$
3. $[c, \neg s, r, \neg w]$
4. $[c, \neg s, \neg r, \neg w]$
5. $[\neg c, \neg s, \neg r, \neg w]$
6. $[\neg c, \neg s, \neg r, \neg w]$
7. $[c, s, r, w]$
8. $[\neg c, s, \neg r, w]$
9. $[c, \neg s, r, w]$
10. $[\neg c, \neg s, r, w]$

$$P(w) = ?$$

[zadanie 24]

Direct sampling — przykład
Wyniki kolejnych próbkowań:
1. $[c, \neg s, r, w]$
2. $[\neg c, s, \neg r, w]$
3. $[c, \neg s, r, \neg w]$
4. $[c, \neg s, \neg r, \neg w]$
5. $[\neg c, \neg s, \neg r, \neg w]$
6. $[\neg c, \neg s, \neg r, \neg w]$
7. $[c, s, r, w]$
8. $[\neg c, s, \neg r, w]$
9. $[c, \neg s, r, w]$
10. $[\neg c, \neg s, r, w]$
$P(w) = ?$
[zadanie 24]

1. Wynik: 6 / 10

Direct sampling — zalety

- Zalety:
  - w krótkim czasie możemy uzyskać aproksymację łącznego rozkładu prawdopodobieństwa
  - możemy je zastosować nawet jeśli nie znamy prawdopodobieństw opisujących model (ale potrafimy symulować proces)

Direct sampling — zalety

- Zalety:
  - w krótkim czasie możemy uzyskać aproksymację łącznego rozkładu prawdopodobieństwa
  - możemy je zastosować nawet jeśli nie znamy prawdopodobieństw opisujących model (ale potrafimy symulować proces)

1. Zaleta próbkowania jest taka, że dokładne wnioskowanie dla całego łącznego rozkładu prawdopodobieństwa ma złożoność wykładniczą, a tutaj możemy przynajmniej w skończonym czasie uzyskać jakieś aproksymacje.

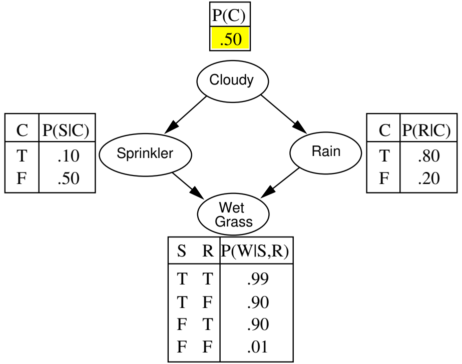
Próbkowanie z odrzucaniem (ang. rejection sampling)

- Pozwala obliczyć rozkład prawdopodobieństwa warunkowego np. $\mathbf{P}(C, R|s, w)$
- Idea:
  1. Wylosuj próbkę
  2. Odrzuć jeśli nie pasuje do wartości zmiennych warunkowych
- Wynik: liczba pasujących / liczba niepasujących

Próbkowanie z odrzucaniem (ang. rejection sampling)

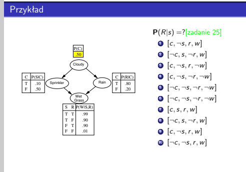
- Pozwala obliczyć rozkład prawdopodobieństwa warunkowego np. $\mathbf{P}(C, R|s, w)$
- Idea:
  1. Wylosuj próbkę
  2. Odrzuć jeśli nie pasuje do wartości zmiennych warunkowych
- Wynik: liczba pasujących / liczba niepasujących

Przykład



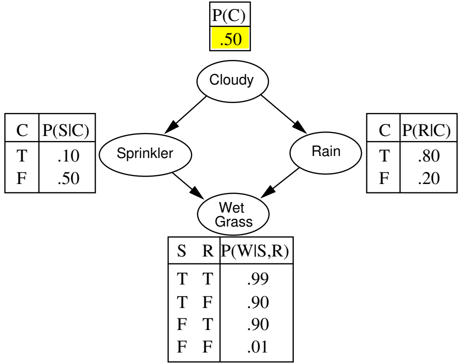
$\mathbf{P}(R|s) = ?$[zadanie 25]

1. $[c, \neg s, r, w]$
2. $[\neg c, s, \neg r, w]$
3. $[c, \neg s, r, \neg w]$
4. $[c, \neg s, \neg r, \neg w]$
5. $[\neg c, \neg s, \neg r, \neg w]$
6. $[\neg c, \neg s, \neg r, \neg w]$
7. $[c, s, r, w]$
8. $[\neg c, s, \neg r, w]$
9. $[c, \neg s, r, w]$
10. $[\neg c, \neg s, r, w]$



1. $\langle \frac{1}{3}, \frac{2}{3} \rangle$
2. Jeśli jakieś prawd. jest bardzo małe (np. prawd. włamania) albo bardzo duże musimy zebrać bardzo dużo próbek, żeby estymacja była dobra.
3. I niestety większość próbek musimy odrzucić

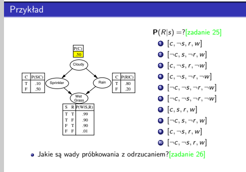
Przykład



$\mathbf{P}(R|s) = ?$[zadanie 25]

1. $[c, \neg s, r, w]$
2. $[\neg c, s, \neg r, w]$
3. $[c, \neg s, r, \neg w]$
4. $[c, \neg s, \neg r, \neg w]$
5. $[\neg c, \neg s, \neg r, \neg w]$
6. $[\neg c, \neg s, \neg r, \neg w]$
7. $[c, s, r, w]$
8. $[\neg c, s, \neg r, w]$
9. $[c, \neg s, r, w]$
10. $[\neg c, \neg s, r, w]$

- Jakie są wady próbkowania z odrzucaniem?[zadanie 26]



1. $\langle \frac{1}{3}, \frac{2}{3} \rangle$
2. Jeśli jakieś prawd. jest bardzo małe (np. prawd. włamania) albo bardzo duże musimy zebrać bardzo dużo próbek, żeby estymacja była dobra.
3. I niestety większość próbek musimy odrzucić

Ważone próbkowanie (ang. likelihood weighting)

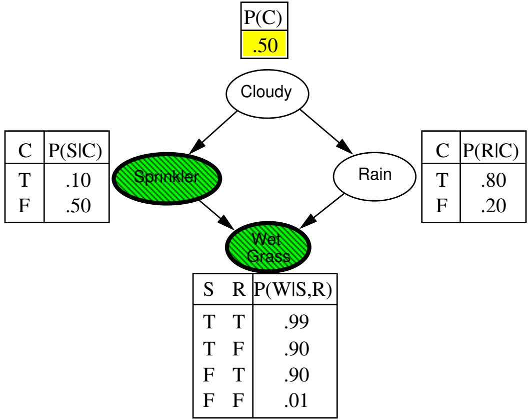
- Ustalamy wartości zmiennych warunkowych $\Rightarrow$ nie trzeba będzie ich odrzucać
- Ważymy próbki odpowiednimi prawdopodobieństwami

Ważone próbkowanie (ang. likelihood weighting)

- Ustalamy wartości zmiennych warunkowych $\Rightarrow$ nie trzeba będzie ich odrzucać
- Ważymy próbki odpowiednimi prawdopodobieństwami

Przykład

$P(c, \neg r | s, w) = ?$



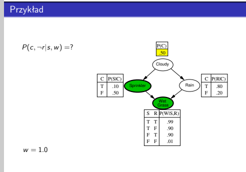
P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

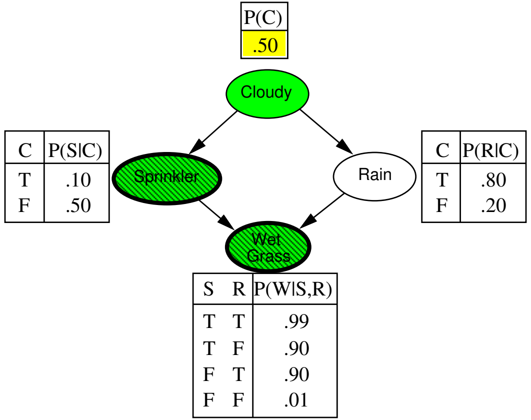
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$w = 1.0$

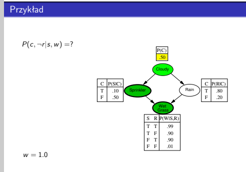


Przykład

$P(c, \neg r | s, w) = ?$

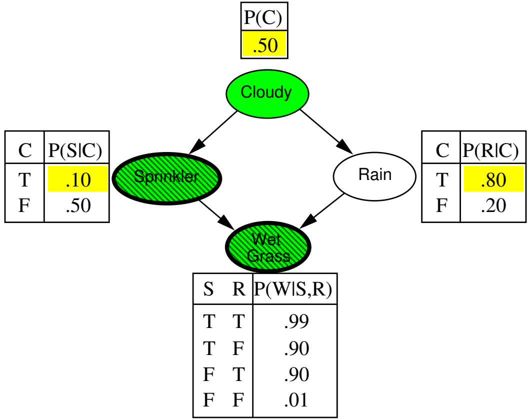


$w = 1.0$



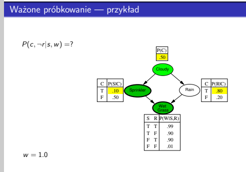
Ważone próbkowanie — przykład

$P(c, \neg r | s, w) = ?$



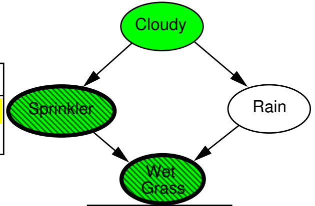
$w = 1.0$

2014-04-03



Ważone próbkowanie — przykład

$P(c, \neg r | s, w) = ?$



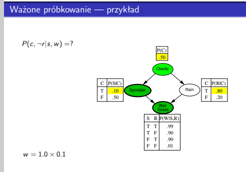
P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

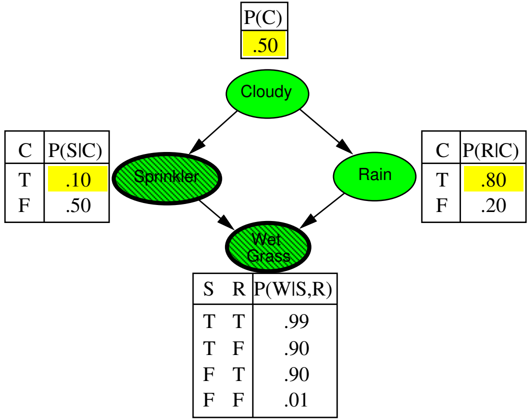
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$w = 1.0 \times 0.1$

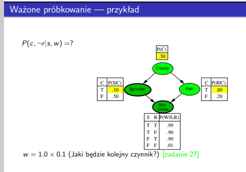


Ważone próbkowanie — przykład

$P(c, \neg r | s, w) = ?$

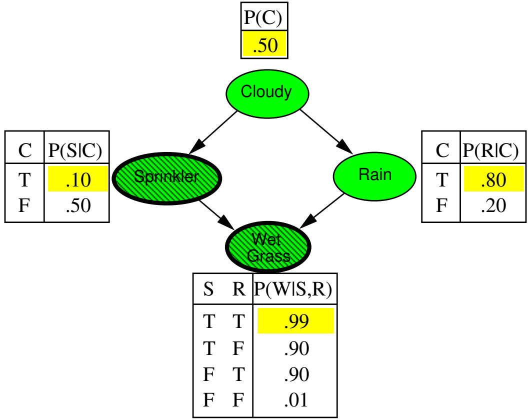


$w = 1.0 \times 0.1$ (Jaki będzie kolejny czynnik?) [zadanie 27]

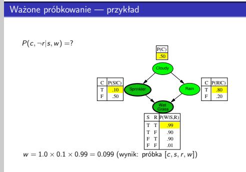


Ważone próbkowanie — przykład

$P(c, \neg r | s, w) = ?$



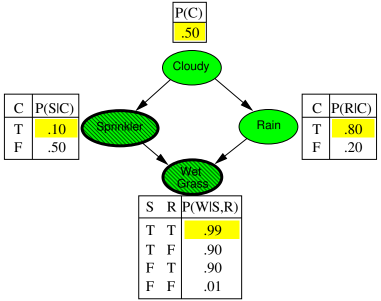
$w = 1.0 \times 0.1 \times 0.99 = 0.099$ (wynik: próbka $[c, s, r, w]$)

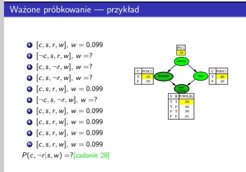


Ważone próbkowanie — przykład

1. $[c, s, r, w],\ w = 0.099$
2. $[\neg c, s, r, w],\ w = ?$
3. $[c, s, \neg r, w],\ w = ?$
4. $[c, s, \neg r, w],\ w = ?$
5. $[c, s, r, w],\ w = 0.099$
6. $[\neg c, s, \neg r, w],\ w = ?$
7. $[c, s, r, w],\ w = 0.099$
8. $[c, s, r, w],\ w = 0.099$
9. $[c, s, r, w],\ w = 0.099$
10. $[c, s, r, w],\ w = 0.099$

$P(c, \neg r | s, w) = ?$[zadanie 28]





1. $0.5 \times 0.99 = 0.495$
2. $0.1 \times 0.90 = 0.09$
3. $0.5 \times 0.90 = 0.45$
4. $(2 \times 0.09)/(6 \times 0.099 + 2 \times 0.09 + 0.495 + 0.45) = 0.1047$

Podsumowanie wnioskowania

- Wnioskowanie **dokładne** (eliminacja zmiennych):
  - w ogólności NP-trudne
  - bardzo wrażliwe na topologię sieci i pytania, które zadajemy
- Wnioskowanie **aproksymacyjne** (próbkowanie):
  - generalnie, nie jest wrażliwe na topologię sieci
  - zbieżność może być powolna, gdy prawd. mają skrajne wartości (0.0 lub 1.0)

Podsumowanie wnioskowania

- Wnioskowanie **dokładne** (eliminacja zmiennych):
  - w ogólności NP-trudne
  - bardzo wrażliwe na topologię sieci i pytania, które zadajemy
- Wnioskowanie **aproksymacyjne** (próbkowanie):
  - generalnie, nie jest wrażliwe na topologię sieci
  - zbieżność może być powolna, gdy prawd. mają skrajne wartości (0.0 lub 1.0)

Podsumowanie wnioskowania

- Wnioskowanie **dokładne** (eliminacja zmiennych):
  - w ogólności NP-trudne
  - bardzo wrażliwe na topologię sieci i pytania, które zadajemy
- Wnioskowanie **aproksymacyjne** (próbkowanie):
  - generalnie, nie jest wrażliwe na topologię sieci
  - zbieżność może być powolna, gdy prawd. mają skrajne wartości (0.0 lub 1.0)

Pominęliśmy:

- Wnioskowanie za pomocą symulacji łańcuchów Markov'a (ang. *Markov chain Monte Carlo*) o nazwie **próbkowanie Gibbs'a.**
  - Pewne zalety względem Likelihood Weighting
- Wiele nowych i efektywniejszych metod

Podsumowanie wnioskowania

- Wnioskowanie **dokładne** (eliminacja zmiennych):
  - w ogólności NP-trudne
  - bardzo wrażliwe na topologię sieci i pytania, które zadajemy
- Wnioskowanie **aproksymacyjne** (próbkowanie):
  - generalnie, nie jest wrażliwe na topologię sieci
  - zbieżność może być powolna, gdy prawd. mają skrajne wartości (0.0 lub 1.0)

Pominęliśmy:

- Wnioskowanie za pomocą symulacji łańcuchów Markov'a (ang. *Markov chain Monte Carlo*) o nazwie **próbkowanie Gibbs'a.**
  - Pewne zalety względem Likelihood Weighting
- Wiele nowych i efektywniejszych metod

Jak konstruować sieci?

- Najgorszy przypadek dla węzła:
  - to $O(2^k)$ liczb, gdzie k = liczba rodziców węzła.
- Zwykle jednak możemy zastosować jeden ze schematów. Rozważymy dwa z nich:
  - **węzeł deterministyczny**
  - **noisy-OR**

Jak konstruować sieci?

- Najgorszy przypadek dla węzła:
  - to $O(2^k)$ liczb, gdzie k = liczba rodziców węzła.
- Zwykle jednak możemy zastosować jeden ze schematów. Rozważymy dwa z nich:
  - **węzeł deterministyczny**
  - **noisy-OR**

Węzeł deterministyczny

- Bez niepewności.
- Węzeł zdeterminowany w pełni przez rodziców (przez wyrażenie logiczne), np.
  - *Polak*, *Niemiec*, *Hiszpan*→*Europejczyk*

P	N	H	P(Europejczyk)
F	F	F	0.0
F	F	T	1.0
F	T	F	1.0
F	T	T	1.0
T	F	F	1.0
T	F	T	1.0
T	T	F	1.0
T	T	T	1.0

Węzeł deterministyczny

- Bez niepewności.
- Węzeł zdeterminowany w pełni przez rodziców (przez wyrażenie logiczne), np.
  - *Polak*, *Niemiec*, *Hiszpan*→*Europejczyk*

P	N	H	P(Europejczyk)
F	F	F	0.0
F	F	T	1.0
F	T	F	1.0
F	T	T	1.0
T	F	F	1.0
T	F	T	1.0
T	T	F	1.0
T	T	T	1.0

Noisy-OR

- Uogólnienie logicznej alternatywy
- Przykład *Przeziębienie*, *Grypa*, *Malaria* $\rightarrow$ *Gorączka*
  - Zakładamy, że znamy wszystkie możliwe przyczyny (lub nowy węzeł *NieznanaPrzyczyna*)
  - Zakładamy, że związek $X \rightarrow Go$ może być w pewnym zakresie „hamowany”
    - **Hamowania są niezależne**, tzn. „hamowanie” $P \rightarrow Go$ jest niezależne od „hamowania” $G \rightarrow Go$.

Noisy-OR

- Uogólnienie logicznej alternatywy
- Przykład *Przeziębienie*, *Grypa*, *Malaria* $\rightarrow$ *Gorączka*
  - Zakładamy, że znamy wszystkie możliwe przyczyny (lub nowy węzeł *NieznanaPrzyczyna*)
  - Zakładamy, że związek $X \rightarrow Go$ może być w pewnym zakresie „hamowany”
    - **Hamowania są niezależne**, tzn. „hamowanie” $P \rightarrow Go$ jest niezależne od „hamowania” $G \rightarrow Go$.

Noisy-OR

- Uogólnienie logicznej alternatywy
- Przykład *Przeziębienie*, *Grypa*, *Malaria* $\rightarrow$ *Gorączka*
    - Zakładamy, że znamy wszystkie możliwe przyczyny (lub nowy węzeł *NieznanaPrzyczyna*)
    - Zakładamy, że związek $X \rightarrow Go$ może być w pewnym zakresie „hamowany"
        - **Hamowania są niezależne**, tzn. „hamowanie" $P \rightarrow Go$ jest niezależne od „hamowania" $G \rightarrow Go$.
- Stąd: gorączka nie występuje $(\neg go) \iff \neg p \wedge \neg g \wedge \neg m$

Noisy-OR

- Uogólnienie logicznej alternatywy
- Przykład *Przeziębienie*, *Grypa*, *Malaria* $\rightarrow$ *Gorączka*
    - Zakładamy, że znamy wszystkie możliwe przyczyny (lub nowy węzeł *NieznanaPrzyczyna*)
    - Zakładamy, że związek $X \rightarrow Go$ może być w pewnym zakresie „hamowany"
        - **Hamowania są niezależne**, tzn. „hamowanie" $P \rightarrow Go$ jest niezależne od „hamowania" $G \rightarrow Go$.
- Stąd: gorączka nie występuje $(\neg go) \iff \neg p \wedge \neg g \wedge \neg m$

Przykład

- Załóżmy, że prawd. „hamowania” są następujące:
  - $P(\neg go|p, \neg g, \neg m) = 0.6$
  - $P(\neg go|\neg p, g, \neg m) = 0.2$
  - $P(\neg go|\neg p, \neg g, m) = 0.1$
- Na tej podstawie budujemy całą tabelkę:

P	G	M	P(*Gorączka*)	P(¬*Gorączka*)
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*0.1
T	F	F	0.4	**0.6**
T	F	T	0.94	0.06=0.6*0.1
T	T	F		?[zadanie 29]
T	T	T		?[zadanie 30]

1. 0.88 i 0.12=0.6*0.2 oraz 0.988 i 0.012=0.6*0.2*0.1

Prawd. całkowite

D_i - pogoda i-tego dnia. $D_i = \{słoneczny, deszczowy\}$

- $P(D_i = s) = 0.9$
- $P(D_{i+1} = s)|D_i = s) = 0.8$
- $P(D_2 = d|D_1 = s) = ?$[zadanie 31]
- $P(D_2 = s|D_1 = d) = 0.6$
- $P(D_2 = d|D_1 = d) = ?$[zadanie 32]
- $P(D_2 = s) = ?$[zadanie 33]
- $P(D_3 = s) = ?$[zadanie 34]
- Narysuj sieć baysowską opisującą ten świat[zadanie 35]

Prawd. całkowite

D_i - pogoda i-tego dnia. $D_i = \{słoneczny, deszczowy\}$

- $P(D_i = s) = 0.9$
- $P(D_{i+1} = s)|D_i = s) = 0.8$
- $P(D_2 = d|D_1 = s) = ?$[zadanie 31]
- $P(D_2 = s|D_1 = d) = 0.6$
- $P(D_2 = d|D_1 = d) = ?$[zadanie 32]
- $P(D_2 = s) = ?$[zadanie 33]
- $P(D_3 = s) = ?$[zadanie 34]
- Narysuj sieć baysowską opisującą ten świat[zadanie 35]

1. 0.2
2. 0.4
3. (prawd. całkowite) = 0.78
4. (prawd. całkowite) = 0.756

Zadania

- Obecność nowotworu: $C = \{c, \neg c\}$.
- Test wykrył raka: $T = \{+, -\}$
- $P(c) = 0.01$
- $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(-|c) =$
- $P(+|\neg c) = 0.2$
- $P(-|\neg c) =$
- $P(+, c) =$
- $P(-, c) =$
- $P(+, \neg c) =$
- $P(-, \neg c) =$

Zadania

- Obecność nowotworu: $C = \{c, \neg c\}$.
- Test wykrył raka: $T = \{+, -\}$
- $P(c) = 0.01$
- $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(-|c) =$
- $P(+|\neg c) = 0.2$
- $P(-|\neg c) =$
- $P(+, c) =$
- $P(-, c) =$
- $P(+, \neg c) =$
- $P(-, \neg c) =$

1. 0.99
2. 0.1
3. 0.8
4. (z definicji) 0.009
5. (z definicji) 0.001
6. (z definicji) 0.198
7. (z definicji) 0.792
8. $P(c|+) = \frac{P(c,+)}{P(+)} = \frac{P(c,+)}{P(+|c)P(c)+P(+|\neg c)P(\neg c)} = \frac{0.9*0.01}{0.9*0.01+0.2*0.99} =$ 0.043
   Zwróćmy uwagę jak mała jest to wartość (w porównaniu do 0.9). Powodem jest prawy składnik licznika (0.2*0.99). Co ciekawe zwiększenie 0.9 nawet do 1.0 niewiele zmienia. Dopiero zwiększenie 0.2 (największe znaczenie!) w dół albo zwiększenie występowania choroby w populacji (0.01) może polepszyć sprawę.

Zadania

- Obecność nowotworu: $C = \{c, \neg c\}$.
- Test wykrył raka: $T = \{+, -\}$
- $P(c) = 0.01$
- $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(-|c) =$
- $P(+|\neg c) = 0.2$
- $P(-|\neg c) =$
- $P(+, c) =$
- $P(-, c) =$
- $P(+, \neg c) =$
- $P(-, \neg c) =$
- $P(c|+) = ?$[zadanie 36]
  - Zwróć uwagę, jak mała jest to wartość w porównaniu z wartością $P(+|c) = 0.9$

Zadania
- Obecność nowotworu: $C = \{c, \neg c\}$.
- Test wykrył raka: $T = \{+, -\}$
- $P(c) = 0.01$
- $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(-|c) =$
- $P(+|\neg c) = 0.2$
- $P(-|\neg c) =$
- $P(+, c) =$
- $P(-, c) =$
- $P(+, \neg c) =$
- $P(-, \neg c) =$
- $P(c|+) = ?$[zadanie 36]
  - Zwróć uwagę, jak mała jest to wartość w porównaniu z wartością $P(+|c) = 0.9$

1. 0.99
2. 0.1
3. 0.8
4. (z definicji) 0.009
5. (z definicji) 0.001
6. (z definicji) 0.198
7. (z definicji) 0.792
8. $P(c|+) = \frac{P(c,+)}{P(+)} = \frac{P(c,+)}{P(+|c)P(c)+P(+|\neg c)P(\neg c)} = \frac{0.9*0.01}{0.9*0.01+0.2*0.99} = 0.043$
   Zwróćmy uwagę jak mała jest to wartość (w porównaniu do 0.9). Powodem jest prawy składnik licznika (0.2*0.99). Co ciekawe zwiększenie 0.9 nawet do 1.0 niewiele zmienia. Dopiero zwiększenie 0.2 (największe znaczenie!) w dół albo zwiększenie występowania choroby w populacji (0.01) może polepszyć sprawę.

Warunkowa niezależność zmiennych

- Obecność nowotworu: C
- Test T wykrył raka $\{+, -\}$. Wynik testu jest losowy. Tutaj przeprowadzamy go dwukrotnie (zmienne losowe T_1, T_2), żeby zwiększyć szansę na prawidłowy wynik.
- Narysuj sieć zawierającą zmienne: C, T_1, T_2[zadanie 37]
- $P(c) = 0.01$, $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(+|\neg c) = 0.2$
- $P(C = c|T_1 = +, T_2 = +) =?$[zadanie 38]

Warunkowa niezależność zmiennych
- Obecność nowotworu: C
- Test T wykrył raka $\{+, -\}$. Wynik testu jest losowy. Tutaj przeprowadzamy go dwukrotnie (zmienne losowe T_1, T_2), żeby zwiększyć szansę na prawidłowy wynik.
- Narysuj sieć zawierającą zmienne: C, T_1, T_2[zadanie 37]
- $P(c) = 0.01$, $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(+|\neg c) = 0.2$
- $P(C = c|T_1 = +, T_2 = +) =?$[zadanie 38]

1. $\alpha P(++|c)P(c) = \alpha P(+|c)P(+|c)P(c) = 0.1698$
   Zmienne T_1 i T_2 są niezależne warunkowo

Warunkowa niezależność zmiennych

- j.w. oblicz:
  - $P(T_2 = +|T_1 = +) =?$[zadanie 39]

Warunkowa niezależność zmiennych

- j.w. oblicz:
  - $P(T_2 = +|T_1 = +) =?$[zadanie 39]

1. Można różnymi sposobami. Tutaj chodzi o pokazanie niezależności warunkowej, więc:
$P(+|+) = P(+|+, c)P(c|+) + P(+|+, \neg c)P(\neg c|+) =$
$P(+|c)P(c|+) + P(+|\neg c)P(\neg c|+) = 0.2301$

Wnioskowanie w sieci BEAMJ

Oblicz prawdopobieństwa w sieci BEAJM:

1. $P(b|j, m)$
2. $P(\neg j|m)$
3. $P(\neg j|a, m)$
4. $P(m|\neg a, j, b)$
5. $P(\neg j, b|m)$

Wnioskowanie w sieci BEAMJ

Oblicz prawdopobieństwa w sieci BEAJM:

1. $P(b|j, m)$
2. $P(\neg j|m)$
3. $P(\neg j|a, m)$
4. $P(m|\neg a, j, b)$
5. $P(\neg j, b|m)$

Ile parametrów potrzeba?

Ile jest parametrów (niestrukturalnych) w sieci?[zadanie 40]

- $A \to B$

Ile parametrów potrzeba?

Ile jest parametrów (niestrukturalnych) w sieci?[zadanie 40]

- $A \to B$

1. $P(a)$ oraz $P(b|a)$ i $P(\neg b|a)$, czyli 3.

Ile parametrów potrzeba?

1. $A \rightarrow B \rightarrow E$
   $A \rightarrow C \rightarrow F$
   $A \rightarrow D \rightarrow F$[zadanie 41]
2.
   - $A, B, C \rightarrow D$, [zadanie 42]
   - $D \rightarrow E, F, G$
   - $C \rightarrow G$

Ile parametrów potrzeba?
$A \rightarrow B \rightarrow E$
$A \rightarrow C \rightarrow F$
$A \rightarrow D \rightarrow F$[zadanie 41]
$A, B, C \rightarrow D$, [zadanie 42]
$D \rightarrow E, F, G$
$C \rightarrow G$

1. 1+2+2+2+2+4=13
2. 1+1+1+8+4+2+2=19

Niezależność warunkowa

- $A \rightarrow B \rightarrow D \rightarrow D \rightarrow G$
- $H \rightarrow G$
- $C \rightarrow B$
- $F \rightarrow E \rightarrow D \rightarrow G$
- Czy:
  - $F \perp A$?[zadanie 43]
  - $F \perp A|D$?[zadanie 44]
  - $F \perp A|G$?[zadanie 45]
  - $F \perp A|H$?[zadanie 46]

Niezależność warunkowa
- $A \rightarrow B \rightarrow D \rightarrow D \rightarrow G$
- $H \rightarrow G$
- $C \rightarrow B$
- $F \rightarrow E \rightarrow D \rightarrow G$
- Czy:
  - $F \perp A$?[zadanie 43]
  - $F \perp A|D$?[zadanie 44]
  - $F \perp A|G$?[zadanie 45]
  - $F \perp A|H$?[zadanie 46]

1. Narysować tę sieć.
2. Tak
3. Nie
4. Nie
5. Tak