

# Úvod do bayesovských sítí

**Jiří Vomlel**

**Tato prezentace je k dispozici na:**

**<http://www.utia.cas.cz/vomlel/>**

## Podmíněná pravděpodobnost

Podmíněná pravděpodobnost  $P(A|B)$  je pravděpodobnostní distribuce veličiny  $A$  splňující vztah  $P(A|B) \cdot P(B) = P(A, B)$ .

Jestliže  $P(B)$  je nenulové pak  $P(A|B) = \frac{P(A,B)}{P(B)}$ .

$$P(\text{male}, \text{long}) = 0.1$$

$$P(\text{male}, \text{short}) = 0.4$$

$$P(\text{female}, \text{long}) = 0.3$$

$$P(\text{female}, \text{short}) = 0.2$$

	<i>long</i>	<i>short</i>
<i>male</i>	$P(m,l) = 0.1$	$P(m,s) = 0.4$
<i>female</i>	$P(f,l) = 0.3$	$P(f,s) = 0.2$

$$P(\text{male}|\text{long}) = \frac{P(\text{male}, \text{long})}{P(\text{long})} = \frac{0.1}{0.4} = 1/4$$

# Bayesův vzorec

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{\sum_A P(A, B)} = \frac{P(B|A) \cdot P(A)}{\sum_A P(B|A) \cdot P(A)}$$

Například mějme dvě veličiny:

- *R - Rain* ... v noci pršelo s hodnotami  $y = yes$  a  $n = no$ .
- *W - Wet grass* ... tráva je mokrá s hodnotami  $y = yes$  a  $n = no$ .

Víme, že:

- jestliže v noci pršelo pak je tráva mokrá s pravděpodobností  $\frac{3}{4}$ ,
- jestliže v noci nepršelo pak je tráva mokrá s pravděpodobností  $\frac{1}{8}$ ,
- pravděpodobnost, že bude v noci pršet je  $\frac{1}{3}$ .

Ráno vidíme, že tráva je mokrá. Jaká je pravděpodobnost, že v noci pršelo?

## Použití Bayesova vzorce

$$\begin{aligned} P(R = y|W = y) &= \frac{P(W = y|R = y) \cdot P(R = y)}{P(W = y|R = y) \cdot P(R = y) + P(W = y|R = n) \cdot P(R = n)} \\ &= \frac{\frac{3}{4} \cdot \frac{1}{3}}{\frac{3}{4} \cdot \frac{1}{3} + \frac{1}{8} \cdot \frac{2}{3}} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{12}} = \frac{\frac{1}{4}}{\frac{4}{12}} = \frac{3}{4} \end{aligned}$$

Pravděpodobnost, že v noci přelo je  $\frac{3}{4}$ .

# Nezávislost



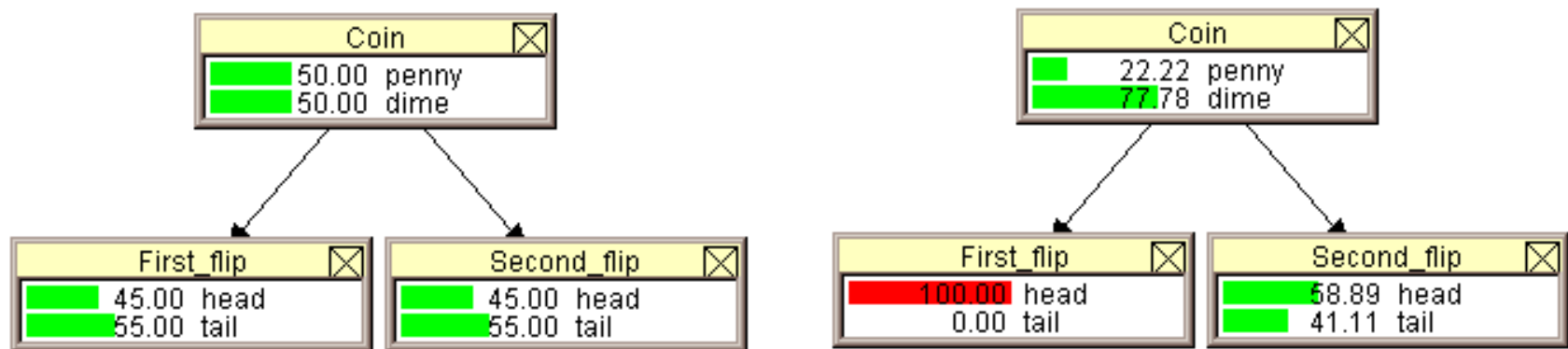
Pravděpodobnost společného výskytu hodnot veličin je rovna součinu pravděpodobností jednotlivých pravděpodobností.

$$P(Dime = head, Penny = head) = P(Dime = head) \cdot P(Penny = head)$$

Též, zjistíme-li hodnotu jedné veličiny, nemá to vliv na hodnotu druhé veličiny.

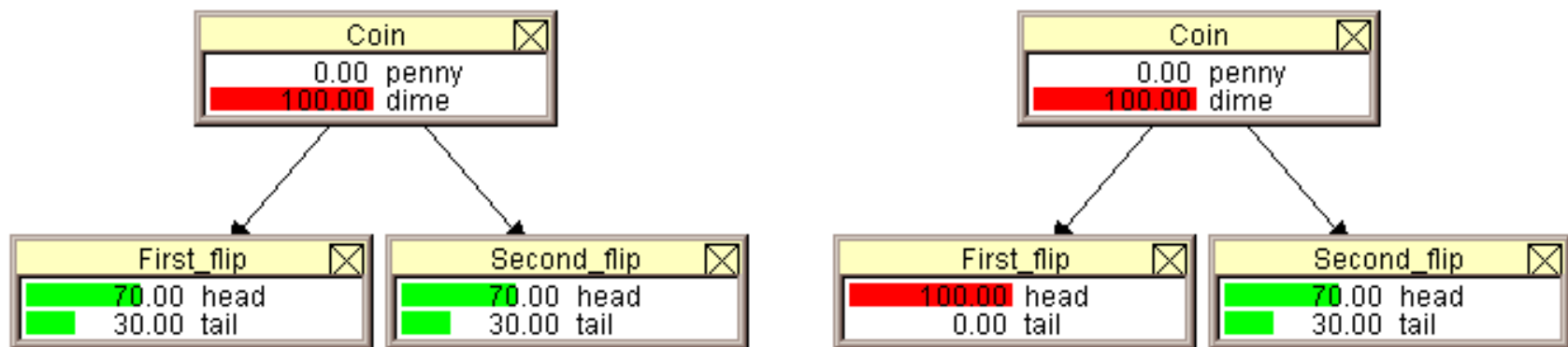
$$P(Dime = head | Penny = head) = P(Dime = head)$$

# Náhodně vybereme jednu minci pro dva hody.



První hod **má vliv** na pravděpodobnost výsledku **druhého** hodu.

Nyní, předpokládejme, že známe vybranou minci.



Jestliže víme,  **která mince**  bude použita pak výsledek  **prvního**  hodů  **nemá vliv**  na pravděpodobnost výsledku  **druhého**  hodů.

# Podmíněná nezávislost

Pravděpodobnost společného výskytu hodnot veličin při dané hodnotě třetí veličiny je rovna součinu pravděpodobností jednotlivých podmíněných pravděpodobností:

$$\begin{aligned} &P(\textit{First\_flip} = \textit{head}, \textit{Second\_flip} = \textit{head} | \textit{Coin} = \textit{dime}) \\ &= P(\textit{First\_flip} = \textit{head} | \textit{Coin} = \textit{dime}) \cdot P(\textit{Second\_flip} = \textit{head} | \textit{Coin} = \textit{dime}) \end{aligned}$$

Jestliže neznáme minci, výsledek **prvního** hodů **má vliv** na pravděpodobnost výsledku **druhého** hodů.

Jestliže známe minci, pak výsledek **prvního** hodů **nemá vliv** na pravděpodobnost výsledku **druhého** hodů.

$$\begin{aligned} &P(\textit{Second\_flip} = \textit{head} | \textit{Coin} = \textit{dime}, \textit{First\_flip} = \textit{head}) \\ &= P(\textit{Second\_flip} = \textit{head} | \textit{Coin} = \textit{dime}) \end{aligned}$$

# Řetězcové pravidlo

S definice podmíněné pravděpodobnosti plyne, že můžeme psát:

$$\begin{aligned}P(A, B, C, D) &= P(A|B, C, D) \cdot P(B, C, D) \\ &= P(A|B, C, D) \cdot P(B|C, D) \cdot P(C, D) \\ &= P(A|B, C, D) \cdot P(B|C, D) \cdot P(C|D) \cdot P(D)\end{aligned}$$

# Proč je Holmesův trávník mokrý?

*Holm* Je Holmesův trávník mokrý?

*Rn* Pršelo v noci?

*Sprnk* Byl Holmesův postřikovač zapnutý?

*Wat* Je Watsonův trávník mokrý?

- Holmesův trávník může být mokrý buď protože pršelo, nebo protože měl zapnutý postřikovač.
- Watsonův trávník může být mokrý protože pršelo. Holmesův postřikovač nemá vliv na Watsonův trávník.
- Déšť nesouvisí s tím, jestli má Holmes zapnutý postřikovač.

Řetězcové pravidlo a podmíněné nezávislosti uvedené výše dávají:

$P(Holm, Wat, Rn, Sprnk)$

$$= P(Holm|Wat, Rn, Sprnk) \cdot P(Wat|Rn, Sprnk) \cdot P(Rn|Sprnk) \cdot P(Sprnk)$$

$$= P(Holm|Rn, Sprnk) \cdot P(Wat|Rn) \cdot P(Rn) \cdot P(Sprnk)$$

# Proč je Holmesův trávník mokrý?

$P(\text{Holm}, \text{Wat}, \text{Rn}, \text{Sprnk})$

$$= P(\text{Holm} | \text{Rn}, \text{Sprnk}) \cdot P(\text{Wat} | \text{Rn}) \cdot P(\text{Rn}) \cdot P(\text{Sprnk})$$

Holmes?				
Edit	Functions	View		
Sprinkler?	yes		no	
Rain?	yes	no	yes	no
yes	1	0.9	1	0
no	0	0.1	0	1

Watson?		
Edit	Functions	View
Rain?	yes	no
yes	1	0.2
no	0	0.8

Sprinkler?	
Edit	View
yes	0.1
no	0.9

Rain?	
Edit	View
yes	0.2
no	0.8

```
graph TD; Rain((Rain?)) --> Watson((Watson?)); Rain((Rain?)) --> Holmes((Holmes?)); Sprinkler((Sprinkler?)) --> Holmes((Holmes?));
```

## Definice bayesovské sítě

- acyklický orientovaný graf (DAG)  $G = (V, E)$
- každý uzel  $i \in V$  odpovídá jedné náhodné veličině  $X_i$  s konečným počtem navzájem disjunktních hodnot  $\mathbb{X}_i$
- $pa(i)$  bude označovat množinu rodičů uzlu  $i$  v grafu  $G$
- ke každému uzlu  $i \in V$  odpovídá podmíněná pravděpodobnostní distribuce  $P(X_i \mid (X_j)_{j \in pa(i)})$
- acyklický orientovaný graf (DAG) reprezentuje podmíněné nezávislostní vztahy mezi veličinami  $(X_i)_{i \in V}$

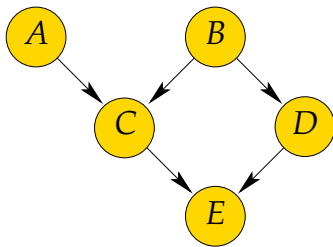
# Podmíněné nezávislost dané grafem - d-separace

Dva uzly  $i$  a  $j$  jsou d-separovány množinou uzlů  $\mathcal{Y}$ , jestliže pro všechny cesty mezi  $i$  a  $j$  platí:

- cesta obsahuje uzel, ve kterém se hrany **nesetkávají** “head-to-head” a který **náleží** do  $\mathcal{Y}$  nebo
- cesta obsahuje uzel, ve kterém se hrany **setkávají** “head-to-head” a ani on, ani žádný jeho následník **nenáleží** do  $\mathcal{Y}$ .

Jestliže  $i$  a  $j$  jsou d-separovány množinou uzlů  $\mathcal{Y}$ , pak veličiny  $X_i$  a  $X_j$  jsou nezávislé dáno  $(X_k)_{k \in \mathcal{Y}}$ .

Příklad:



$$A \perp\!\!\!\perp D$$

$$A \not\perp\!\!\!\perp D \mid E$$

$$C \not\perp\!\!\!\perp D$$

$$C \perp\!\!\!\perp D \mid B$$

# Pravděpodobnostní distribuce bayesovské sítě

Použijeme-li **řetězcové pravidlo** dostaneme:

$$P((X_i)_{i \in V}) = \prod_{i \in V} P(X_i \mid X_{i-1}, \dots, X_1)$$

Použijeme-li podmíněné nezávislosti z grafu pak dostaneme

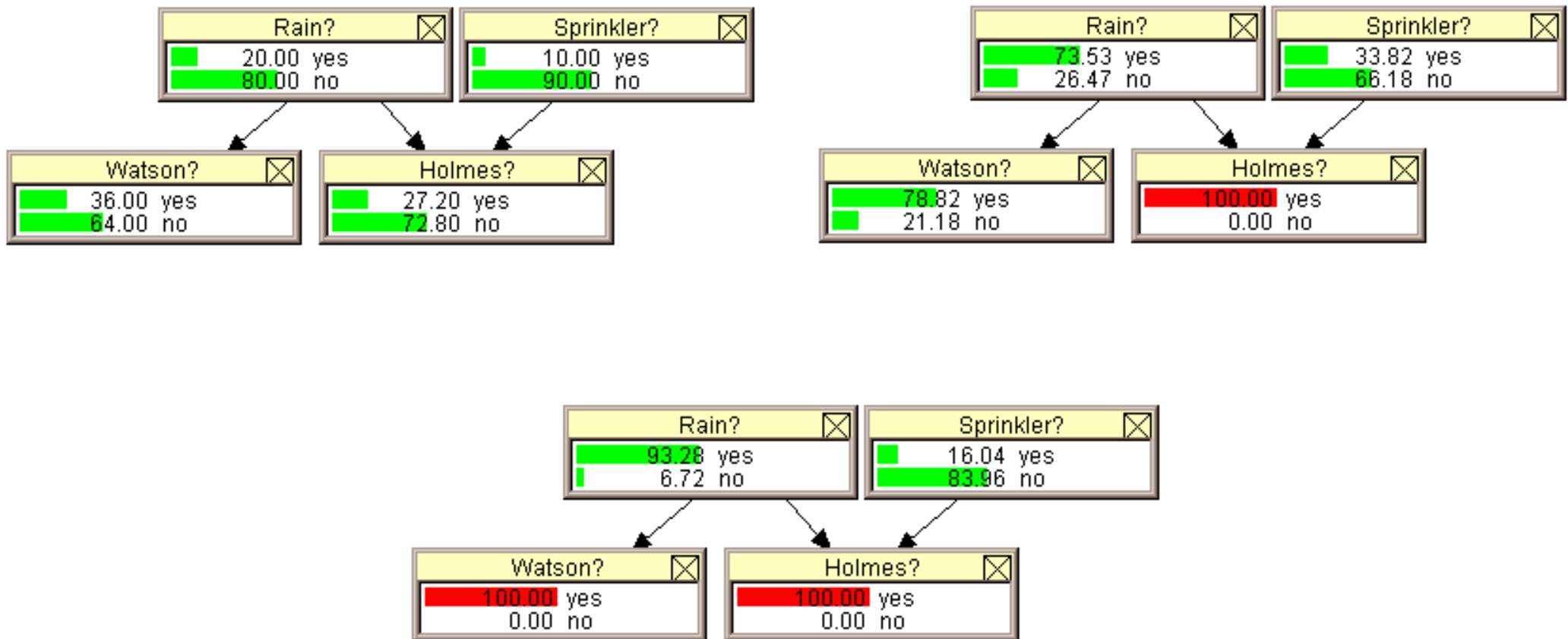
$$P((X_i)_{i \in V}) = \prod_{i \in V} P(X_i \mid (X_j)_{j \in pa(i)})$$

- pravděpodobnostní distribuci representovanou **bayesovskou sítí**.

Pro tuto distribuci platí:

- splňuje všechny podmíněné nezávislosti z grafu a
- její podmíněné pravděpodobnostní distribuce odpovídají  $P(X_i \mid (X_j)_{j \in pa(i)}), i \in V$ .

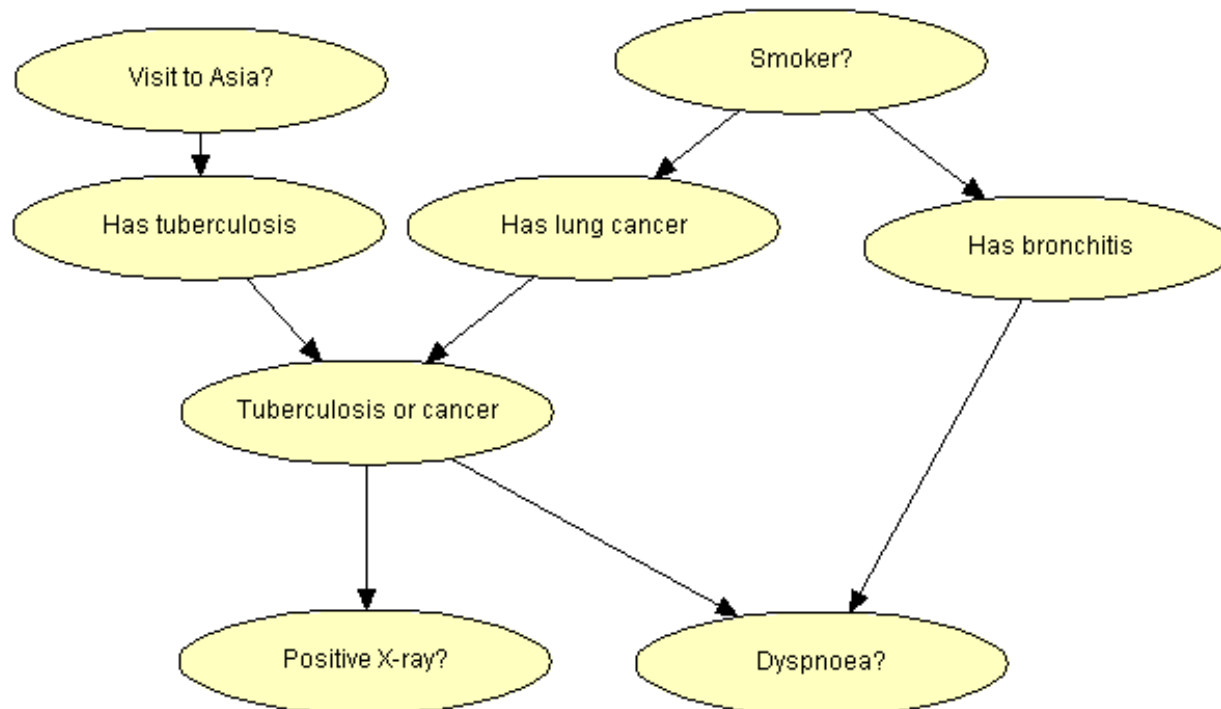
# Byl Holmesův postřikovač zapnutý?



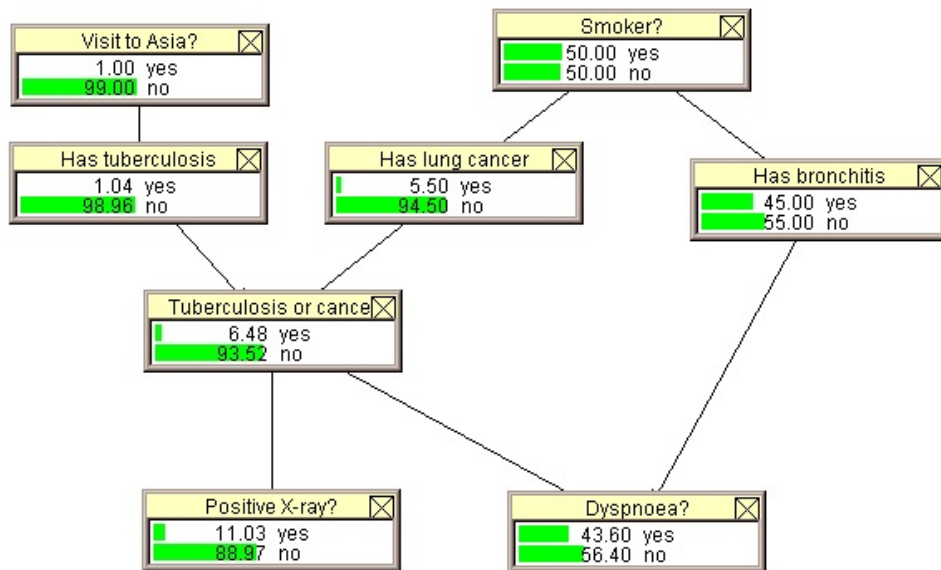
# Zjednodušený diagnostický příklad

Vyšetřujeme pacienta.

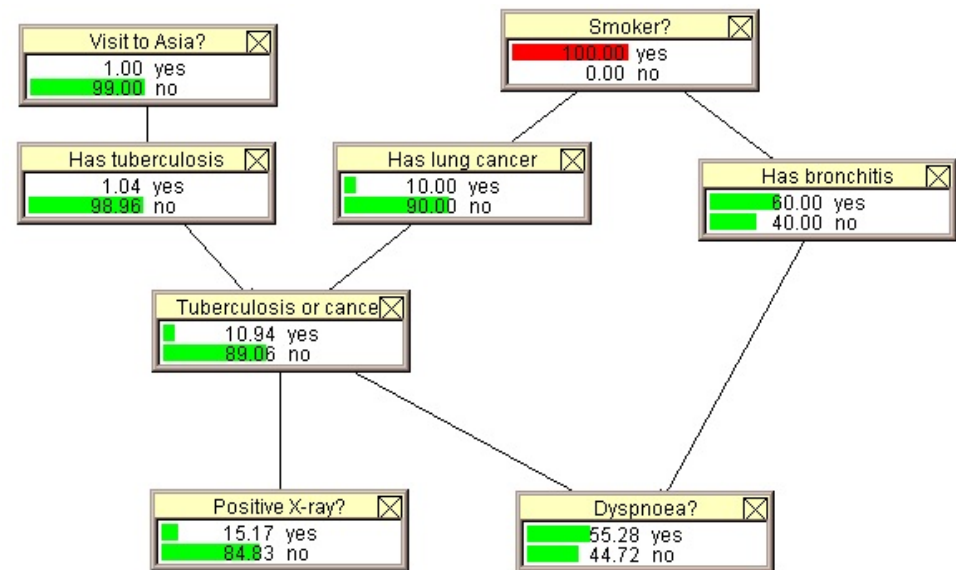
Možná diagnóza je: tuberkulóza, rakovina plic, nebo zánět průdušek.



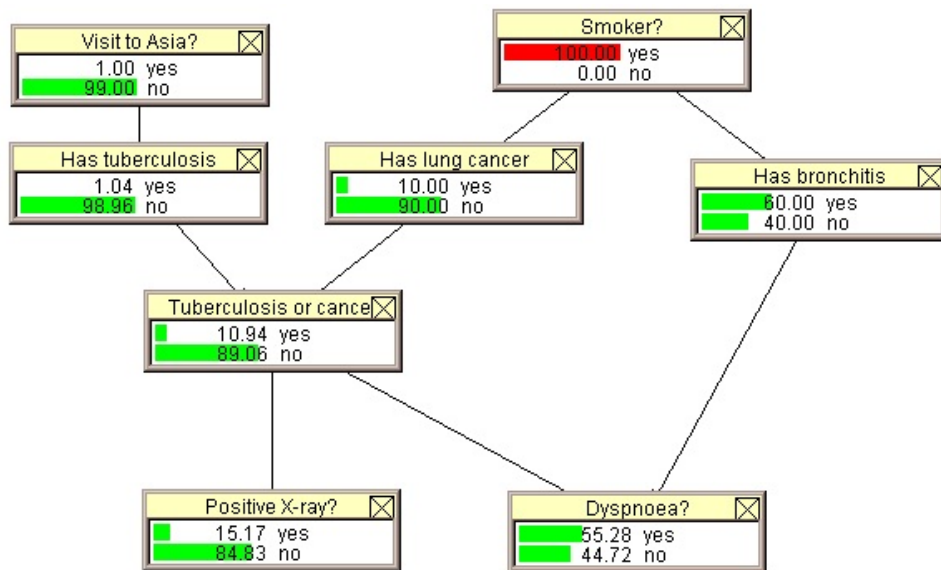
Zatím nevíme o pacientovi nic.



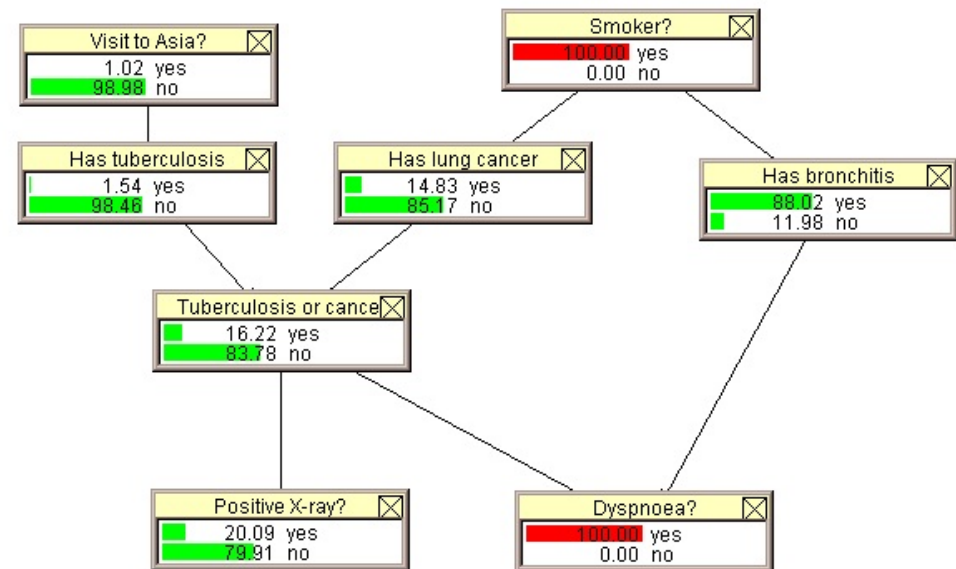
Pacient je kuřák.



Pacient je kuřák

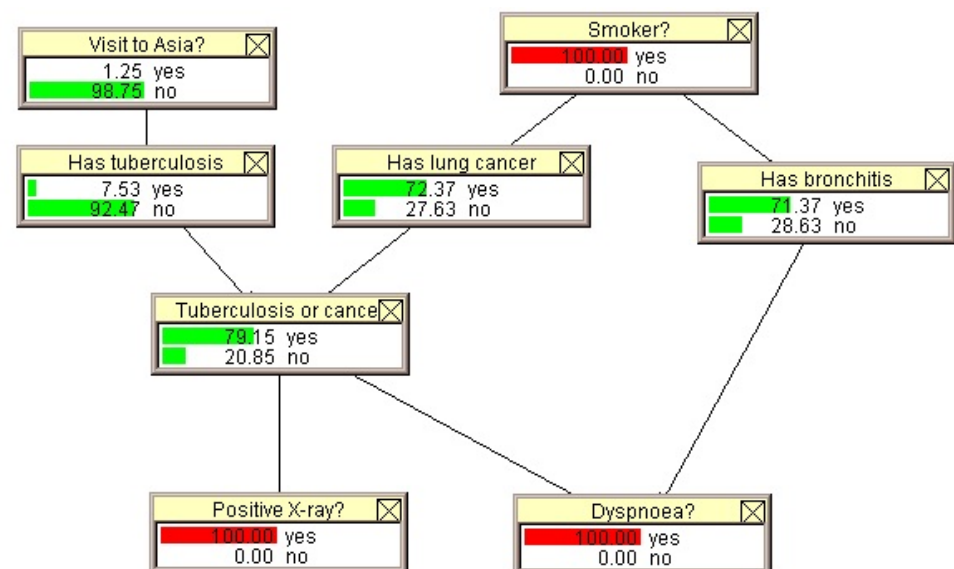
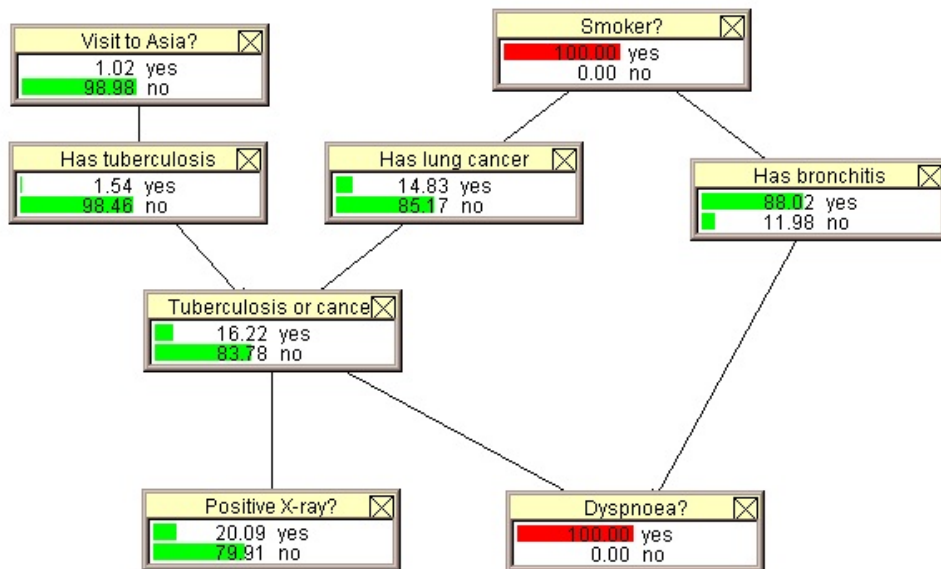


... a navíc si stěžuje na dušnost



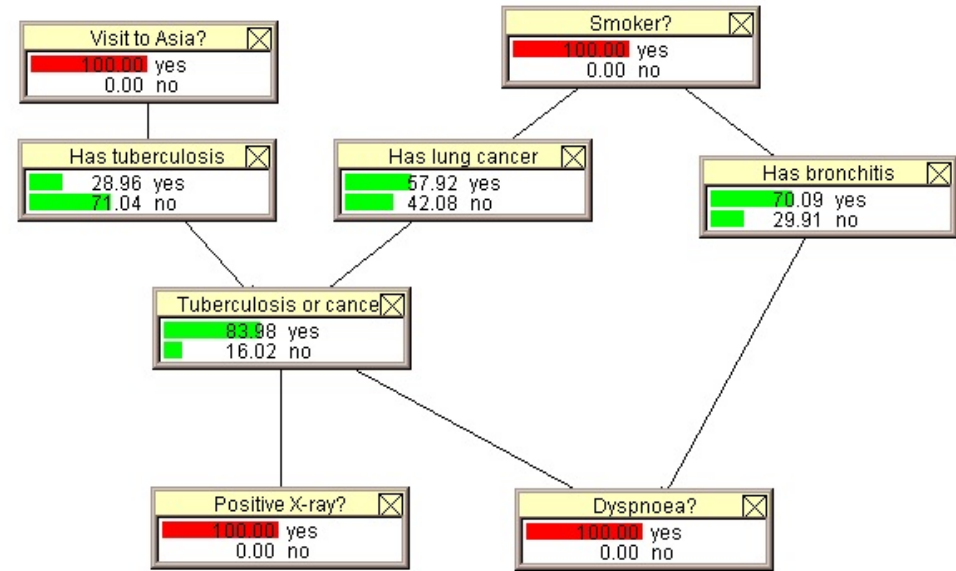
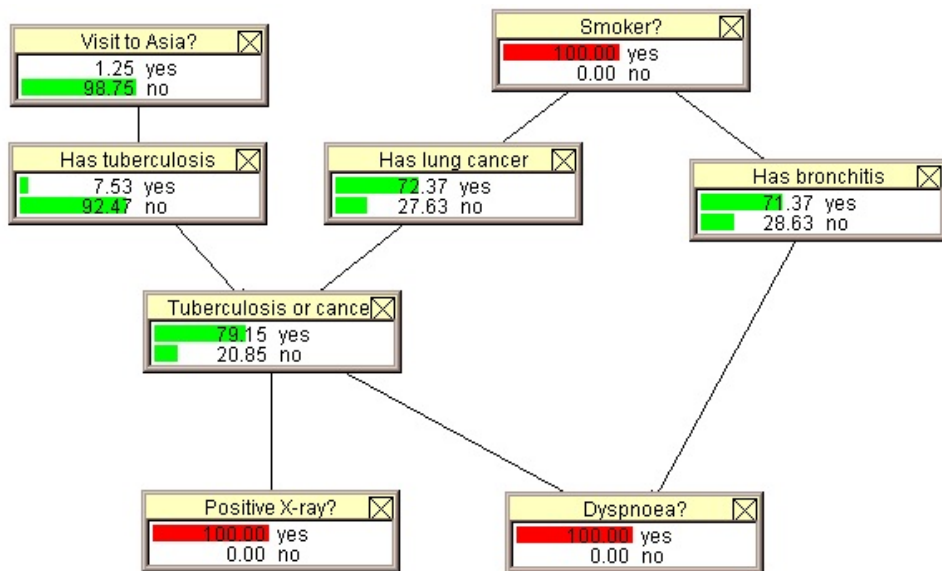
Pacient je kuřák, stěžuje si na dušnost

... a navíc jeho RTG je pozitivní



Pacient je kuřák, stěžuje si na dušnost,  
jeho RTG je pozitivní

... a navíc byl v Asii

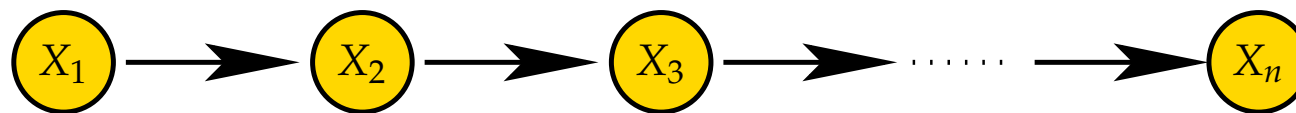


# Výhoda reprezentace bayesovskou sítí

Předpokládejme, že máme problém, který budeme modelovat pomocí  $n$  veličin a každá veličina může nabývat dvou hodnot.

Použijeme-li reprezentaci pomocí jedné tabulky potřebujeme pro uložení v paměti počítače distribuce  $2^n - 1$  hodnot.

Předpokládejme, bayesovskou síť mající též  $n$  veličin nabývajících dvou hodnot s grafem následující struktury:



Pro její uložení v paměti počítače potřebujeme  $1 + (n - 1) \cdot 2 = 2n - 1$  hodnot.

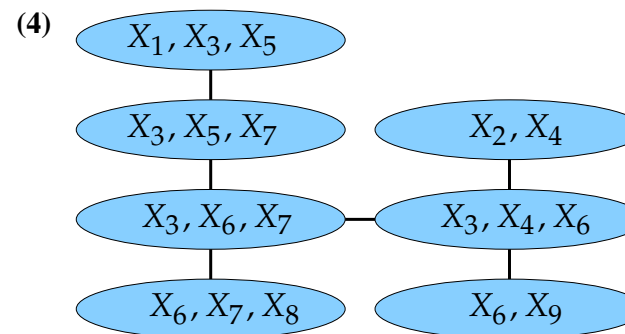
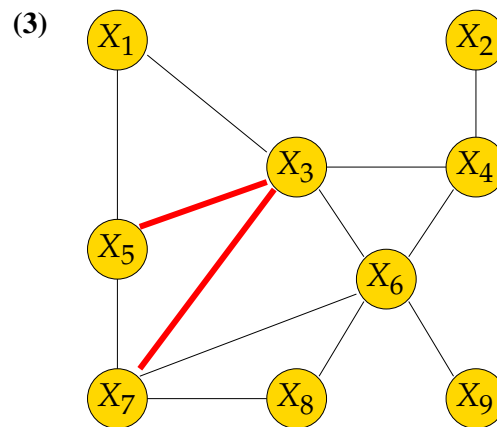
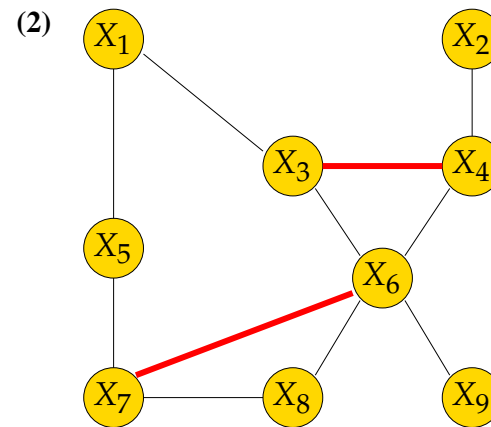
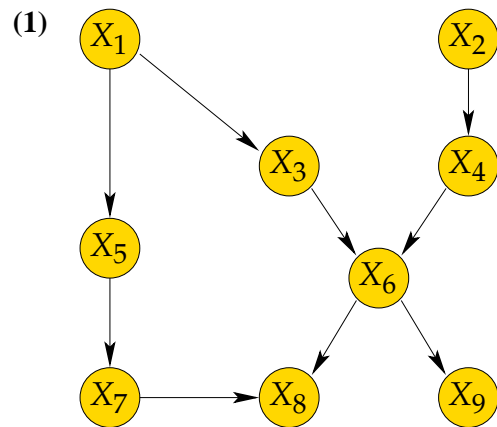
# Výhoda reprezentace bayesovskou sítí

$n$	$2^n - 1$	$2n - 1$
2	3	3
3	7	5
4	15	7
10	1023	19
100	$1.27 \cdot 10^{30}$	199
1000	$1.07 \cdot 10^{301}$	1999

# Typické použití bayesovských sítí

- pro **modelování** a **vysvětlení** chování, problémů v různých oblastech - např. **model chování vody v krajině**,
- pro **přepočtení pravděpodobností** hodnot určitých veličin když jsme pozorovali některé jiné veličiny, t.j. počítání podmíněných pravděpodobností, např.  $P(X_{23} | X_{17} = \text{yes}, X_{54} = \text{no})$  - např. **určení diagnózy při pozorovaných příznacích**.
- pro nalezení **nejpravděpodobnějších konfigurací** proměných - např. **při automatickém rozpoznávání řeči**, nebo **při dekódování zakódovaných zpráv**,
- pro podporu **rozhodování** při nejisté informaci - použití **teorie maximalizace očekávaného užitku**,
- pro nalezení dobrých **strategií** pro řešení problémů v oblastech z nejistotou - např. **technická diagnostika laserových tiskáren**, nebo **návrh adaptivních testů**.

# Výpočty ve stromech spojení (junction tree)



# Z historie bayesovských sítí - 1

Bayesovské sítě patří mezi pravděpodobnostní grafické modely.

- První použití modelů podobných grafickým modelům ve statistické mechanice (Gibbs, 1902),  
genetice (Wright, 1921),  
analýze kontingenčních tabulek (Barlett, 1935).
- Skutečné rošíření myšlenky použití grafických modelů až v osmdesátých letech 20.století:  
Pearl (1982),  
Spiegelhalter a Kill-Jones (1984),  
Perez a Jiroušek (1985),  
Lauritzen a Spiegelhalter (1988)

## Z historie bayesovských sítí - 2

- Vychází základní monografie o grafických modelech a bayesovských sítích:

Pearl (1988), Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,

Neapolitan (1990), Probabilistic reasoning in expert systems: theory and algorithms,

Hájek, Havránek, Jiroušek (1991), Uncertain Information Processing in Expert Systems,

Lauritzen (1996), Graphical models,

Jensen (1996), An introduction to Bayesian networks.

## Z historie bayesovských sítí - 3

- Objevují se aplikace bayesovských sítí v různých oblastech:
  - Munin (1989) - Bayesian network for the median nerve
  - Pathfinder (1990) - lymph-node pathology,
  - Decision theoretic troubleshooting (1994) - technická diagnostika zařízení.
- Objevují se první verze software umožňující práci s bayesovskými sítěmi:
  - Hugin (1989) <http://www.hugin.com>,
  - Netica (1995) <http://www.norsys.com>,