

Při práci s Bayesovskou sítí musíme:

- a) navrhnout strukturu sítě (tj. závislosti)
 - b) -- výchozí pravděpodobnosti
 - c) provést příslušné výpočty
- } výstavba modelu

Návrh struktury Bayesovské sítě

Příklad: Farmař pěstuje brávy na mléko. Mléko může být infikované, a tak si provádí test, který vyjde buď negativní nebo pozitivní. Test ale není stoprocentní, někdy dává pozitivní výsledek pro zdravé mléko a naopak.

- 1) určíme proměnné a jejich stavy

I ... infikované mléko ... stavy: ano, ne

T ... výsledek testu ... stavy: poz, neg

- 2) určíme směr závislosti

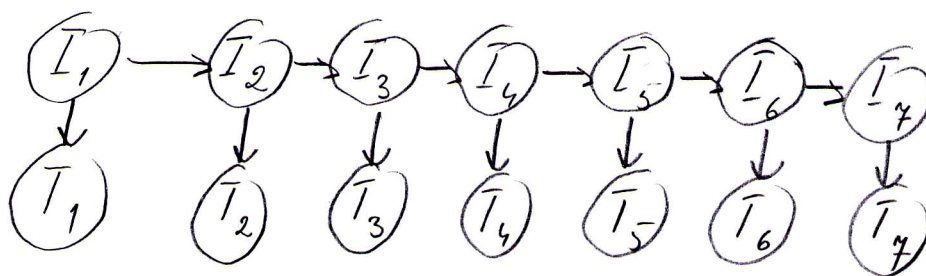


infekce v mléce ovlivní výsledek testu, ne naopak

- 3) vezmeme v úvahu, že test se provádí denně a že jsou vždy k dispozici výsledky za uplynulý týden

- 4) předpokládáme, že kráva, která má jeden den infikované mléko, ho může mít infikované i následující den (je-li nějak nemocná apod.)

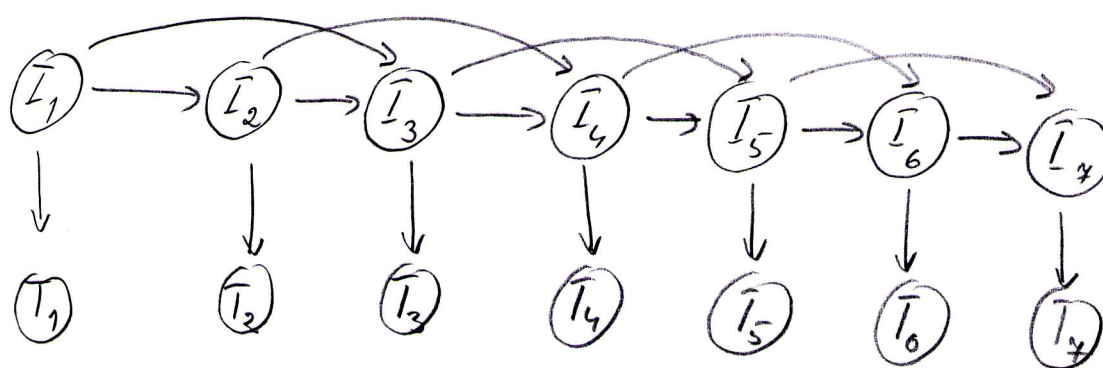
Takže máme model:



- 5) V tomto sériovém propojení $I_1, \dots, I_7$ je ale proměnná I_{i-1} d-separována od I_{i+1} při daném I_i .

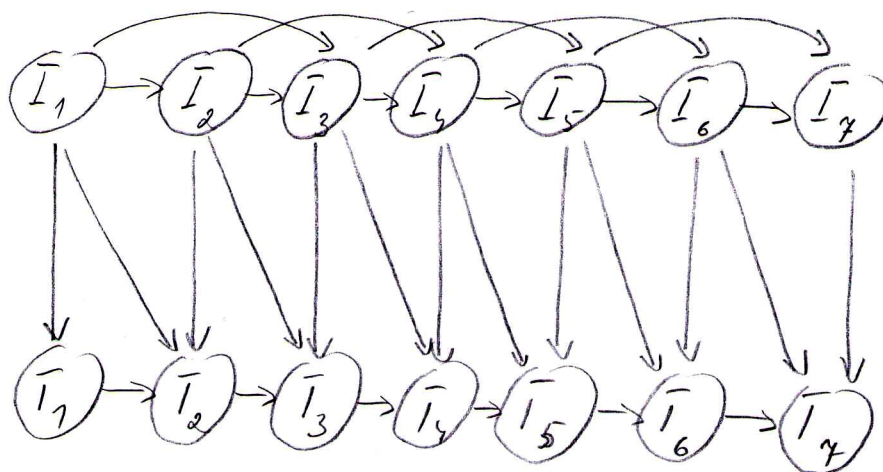
To ale není přirozené např. v situaci, kdy infekce je způsobena nemocí krávy - je-li kráva nemocná dnes a včera byla zdravá, pravděpodobně bude nemocná i zítra nebo pozítří; je-li ale nemocná už čtyři dny, tak zítra už by mohla být zase zdravá ... apod.

Takže to upřesníme třeba takto:



- 6) Další možnost je předpokládat závislost mezi výsledky testu, která je jednak zprostředkována příslušnými vzorky mléka (mezi kterými už jsme závislost zavedli), jednak závislostí korektnosti testu (bude test, který včera fungoval korektně, korektní i dnes?)

Další možný model by tedy byl:

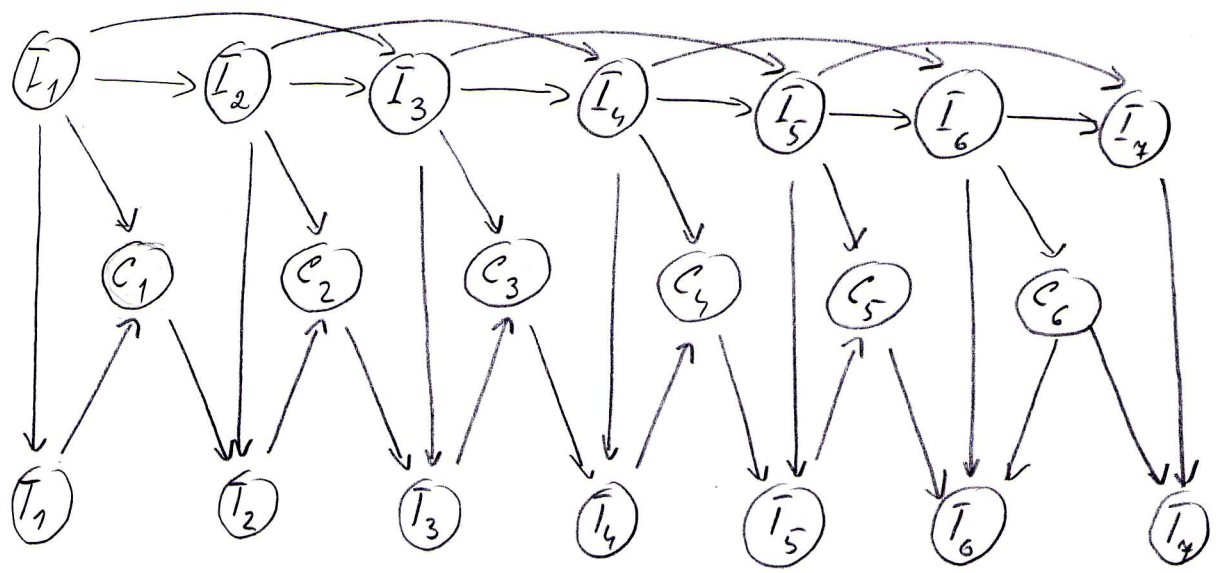


Překvapivě jednodušší pro počítání je ale zavést ještě jednu zprostředkující proměnnou

4) Takže zavedeme ještě jednu proměnnou

C ... borektnost testu ... stavy: ano, ne

Výsledný model může vypadat takto:



Návrh výchozích pravděpodobností

Co budeme počítat - cílem je na základě testu rozhodnout, zda mléko je infikované, tedy $P(I_i | T_i)$

Co k tomu budeme potřebovat?

Předně $P(I)$, $P(T|I)$ - to stačí pro nejjednodušší model bez paměti

Pak $P(I_{i+1} | I_i, I_{i-1})$
 $P(C_i | I_i, T_i)$
 $P(T_i | I_i, C_{i-1})$ } pro poslední model

$P(T|I)$ se stanoví buď podle parametrů testu dodaných jeho výrobcem nebo statisticky na základě dlouhodobých zkušeností s jeho užíváním,

např. $P(\bar{I} = \text{poz} | I = \text{ne}) = P(\bar{I} = \text{neg} | I = \text{ano}) = 0.01 \%$

$P(I)$ — stanovení apriorních pravděpodobností je v Bayesovské analýze to nejosemetnější — různé se odhadují, např. na základě relativních četností, podle principu neurčitosti (tj. všechny možné výsledky mají stejnou pravděpodobnost) apod.

Pro výpočet jsou důležité, jejíž špatná volba může mít značný vliv na výsledky.

V našem případě princip neurčitosti zavrhneme, znamená by, že $P(\bar{I} = \text{ano}) = P(I = \text{ne}) = \frac{1}{2}$

Jinak: π = denní frekvence infikovaného mléka
pro každou krávu na farmě (necht' jich je 50)

Potíž je, že mléko od všech krav se slévá dohromady a nedá se zjistit, pokud mléko bylo infikované, která kráva to způsobila — budeme tedy předpokládat, že π bude pro všechny krávy stejné, a že "krávy jsou nezávislé"

$$\pi = P(I = \text{ano}), \quad 1 - \pi = P(\bar{I} = \text{ne})$$

Necht' infekční mléko (potvrzene přesnými testy veterinařů) se objeví v průměru jednou za měsíc,

%

(20a)

Podle Bayesova vzorce $P(I|T) = \frac{P(T|I)P(I)}{P(T)}$

takže $P(\bar{I} = \text{ano} | T = \text{poz}) = \frac{P(T = \text{poz} | \bar{I} = \text{ano})P(\bar{I} = \text{ano})}{P(T = \text{poz})}$

$$= \frac{\overbrace{P(T = \text{poz} | \bar{I} = \text{ano})}^{0.99} \overbrace{P(\bar{I} = \text{ano})}^{\pi}}{P(T = \text{poz} | \bar{I} = \text{ano})P(\bar{I} = \text{ano}) + P(T = \text{poz} | \bar{I} = \text{ne})P(\bar{I} = \text{ne})}$$

$$= \frac{0.99 \times 0.0004}{0.99 \times 0.0004 + 0.01 \times 0.9993} = \frac{0.000396}{0.010686} = 0.0648$$

$$P(\bar{I} = \text{ne} | T = \text{poz}) = \frac{P(T = \text{poz} | \bar{I} = \text{ne})P(\bar{I} = \text{ne})}{P(T = \text{poz} | \bar{I} = \text{ne})P(\bar{I} = \text{ne}) + P(T = \text{poz} | \bar{I} = \text{ano})P(\bar{I} = \text{ano})}$$

$$= \frac{0.01(1-\pi)}{0.01(1-\pi) + 0.99\pi} = \frac{0.009993}{0.010686} = 0.9351$$

$$P(\bar{I} = \text{ano} | T = \text{neg}) = \frac{P(T = \text{neg} | \bar{I} = \text{ano})P(\bar{I} = \text{ano})}{P(T = \text{neg} | \bar{I} = \text{ano})P(\bar{I} = \text{ano}) + P(T = \text{neg} | \bar{I} = \text{ne})P(\bar{I} = \text{ne})}$$

$$= \frac{0.01\pi}{0.01\pi + 0.99(1-\pi)} = \frac{0.000007}{0.989314} = 0.00000707$$

$$P(\bar{I} = \text{ne} | T = \text{neg}) = \frac{P(T = \text{neg} | \bar{I} = \text{ne})P(\bar{I} = \text{ne})}{P(T = \text{neg} | \bar{I} = \text{ne})P(\bar{I} = \text{ne}) + P(T = \text{neg} | \bar{I} = \text{ano})P(\bar{I} = \text{ano})}$$

$$= \frac{0.99(1-\pi)}{0.99(1-\pi) + 0.01\pi} = \frac{0.989307}{0.989314} = 0.999999$$

tj. 29 dní ze 30 žádná kráva nebyla infikována

$$\text{máme: } (1 - \pi)^{50} = \frac{29}{30}$$

$$\text{z toho } \pi \approx 0,0007$$

Zbývající podmíněné pravděpodobnosti se už musí stanovit na základě zkušenosti.

Např. vyjdeme ze zkušenosti, že:

riziko náhlého onemocnění zdravé krávy je 0,0002

nemoc trvá vždy alespoň 2 dny

šance, že po 2 dnech bude kráva zdravá, je 0,4

pak:

$$P(\bar{I}_{i+1} = \text{ano} \mid \bar{I}_i = \text{ano}, \bar{I}_{i-1} = \text{ano}) = 1 - \text{pst uzdravení po 2 dnech} = 1 - 0,4 = 0,6$$

$$P(\bar{I}_{i+1} = \text{ano} \mid \bar{I}_i = \text{ano}, \bar{I}_{i-1} = \text{ne}) = 1, \text{ protože nemoc trvá nejmeně 2 dny}$$

$$\begin{aligned} P(I_{i+1} = \text{ano} \mid \bar{I}_i = \text{ne}, \bar{I}_{i-1} = \text{ano}) &= \\ &= P(\bar{I}_{i+1} = \text{ano} \mid \bar{I}_i = \text{ne}, \bar{I}_{i-1} = \text{ne}) = \\ &\text{pst náhlého onemocnění} = 0,0002 \end{aligned}$$

$$P(\bar{I}_{i+1} = ne \mid \bar{I}_i = ano, \bar{I}_{i-1} = ano) = 0.4$$

$$P(\bar{I}_{i+1} = ne \mid \bar{I}_i = ano, \bar{I}_{i-1} = ne) = 0$$

$$P(\bar{I}_{i+1} = ne \mid \bar{I}_i = ne, \bar{I}_{i-1} = ano) = 1 - 0.0002 = 0.9998$$

$$= P(\bar{I}_{i+1} = ne \mid \bar{I}_i = ne, \bar{I}_{i-1} = ne)$$

Pokud jde o korektnost testu, vi se, že korektní test má šanci 0.999 být korektní i na sledující dep, nekorektní test bude nekorektní i zítra s psti 0.9.

Takže xřejmē:

$$P(C = ano \mid \bar{T}_i = poz, \bar{I}_i = ano) = 1$$

$$P(C = ano \mid \bar{T}_i = poz, \bar{I}_i = ne) = 0$$

$$P(C = ano \mid \bar{T}_i = neg, \bar{I}_i = ano) = 0$$

$$P(C = ano \mid \bar{T}_i = neg, \bar{I}_i = ne) = 1$$

$$\text{analogicky } P(C = ne \mid \bar{T}_i, \bar{I}_i)$$

a

$$P(\bar{T}_i = poz \mid \bar{I}_i = ano, C_{i-1} = ano) = 0.999$$

$$P(\bar{T}_i = poz \mid \bar{I}_i = ano, C_{i-1} = ne) = 0.1$$

$$P(\bar{T}_i = poz \mid \bar{I}_i = ne, C_{i-1} = ano) = 0.001$$

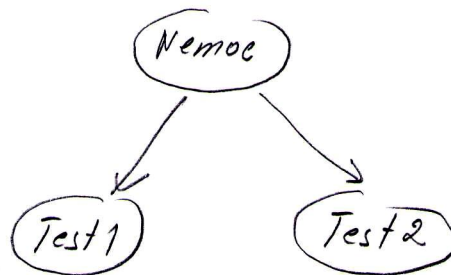
$$P(\bar{T}_i = poz \mid \bar{I}_i = ne, C_{i-1} = ne) = 0.9$$

Takže máme všechny výchozí pravděpodobnosti a výpočty
se pak už provedou pomocí příslušného softwaru -
přehled je např. na

<http://www.es.ube.ca/~murphy6/software/bnsoft.html>
~~replakaj otter 2~~

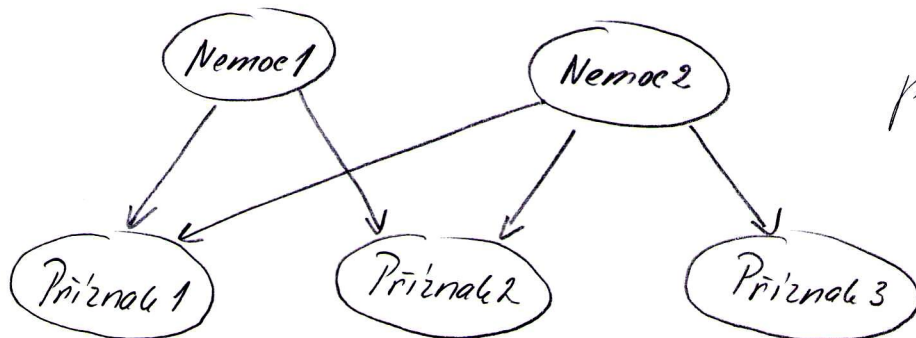
Jiné typické příklady použití Bayesovských sítí - viz Jensen

Diagnostické modely typu



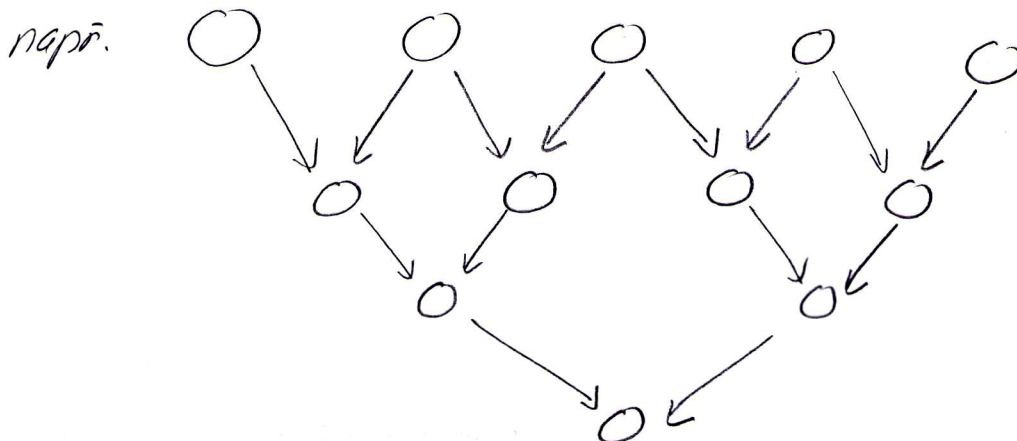
podle výsledků testů se
usuzuje na nemoc

nebo



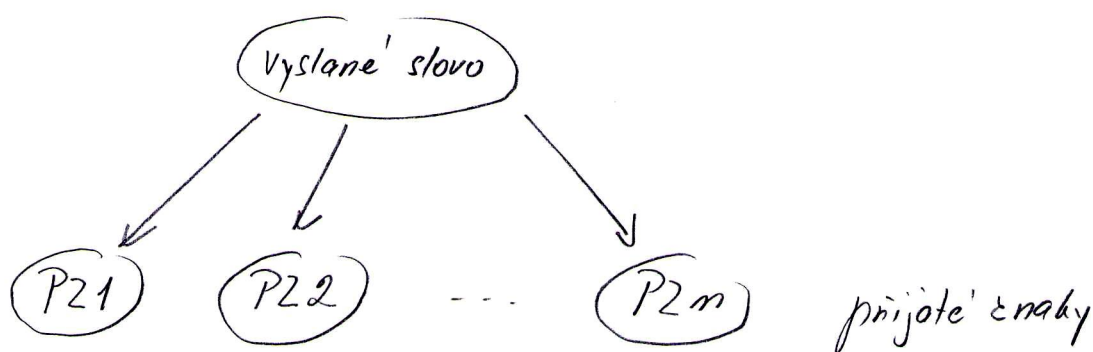
podle příznaků
se usuzuje
na nemoc

Rodokmeny (případně plemenitba zvířat, krížení rostlin)



u posledního jedince byla zjištěna nějaká dědičná choroba
patří se v genotypu předků, kdo ji způsobil

Přenosy znaků (slov, dat) přenosovým kanálem



z přijatých znaků (a ze znalosti kvality přenosového
zařízení) se usuzuje, co bylo vlastně vysláno

Modely pravděpodobnosti v baretních hrách (poker) -

tomu nerozumím

ale jde o to odhadnout, jaké karty má protivník

Obecná pravidla budování modelu

- 1) určit typy proměnných: hypotézy (nemůžu je přímo pozorovat,
ale jsou to právě ty, které mě zajímají)
informační - ty pozoruju
zprostředkující - vkládají se uměle za nějakým účelem
(zajišťují správnou závislost mezi proměnnými,
zjednodušují výpočty apod.)

2) stanovit směr kauzality

3) učení výchozích pravděpodobností:

na základě obecně známých pravděpodobností

(třeba že při rozdávání dobře rozmíchaných karet je každá karta stejně pravděpodobná)

na základě frekvencí (chlapců se rodí přibližně stejně jako dívek)

na základě subjektivních zkušeností a názorů

(při angíně mě vždycky bolí v krku ale malolbky mám horečku)

na základě rady experta (expert může být v modelu

zahrnut jako eprostředkující proměnná)