

Rețele bayesiene

IA 2022/2023

Introducere

Inferență

Independență

Teorema lui Bayes

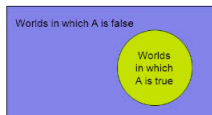
Rețele bayesiene

Inferență în rețele bayesiene

Probabilități

Probabilitatea asociată unei propoziții este suma probabilităților lumilor în care aceasta este adevărată.

$$P(\phi) = \sum_{w \in \phi} P(w), \quad \phi \text{ propoziție}$$



$P(A)$ is the area of the oval

Exemplu: $P(\text{Total} = 11) = P((5, 6)) + P((6, 5)) = 1/36 + 1/36 = 1/18$.

Presupunem $P(\text{doubles}) = 1/4$.

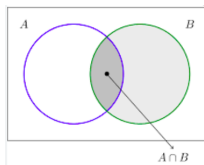
Probabilități necondiționate (*prior*)

Probabilități condiționate

Probabilități condiționate (*posterior*) $P(A|B)$ este fracțiunea de lumi posibile în care B este adevărată și atunci și A este adevărată

- probabilitatea lui A , dat fiind B

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$



Exemplu: $P(\text{doubles} | \text{Die}_1 = 5) = \frac{P(\text{doubles} \wedge \text{Die}_1 = 5)}{P(\text{Die}_1 = 5)}$.

Regula produs: $P(a \wedge b) = P(a|b)P(b)$

- Distribuție de probabilitate a unei variabile aleatoare

$$P(\text{Weather} = \text{sunny}) = 0.6$$

$$P(\text{Weather} = \text{rain}) = 0.1$$

$$P(\text{Weather} = \text{cloudy}) = 0.29$$

$$P(\text{Weather} = \text{snow}) = 0.01$$

$$\mathbf{P}(\text{Weather}) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$$

- Funcția densitate de probabilitate pentru o var. aleatoare continuă

$P(\text{NoonTemp} = x) = \text{Uniform}_{[18C, 26C]}(x)$ temperatura la prânz e distribuită uniform între 18C și 26C

$$P(x) = \lim_{dx \rightarrow 0} P(x \leq X \leq x + dx) / dx$$

Distribuția comună de probabilitate

$\mathbf{P}(\textit{Weather}, \textit{Cavity})$: probabilitățile tuturor combinațiilor de valori ale var. *Weather* și *Cavity*.

(o tabela 4×2)

Putem utiliza $\mathbf{P}(\textit{Weather}, \textit{Cavity}) = \mathbf{P}(\textit{Weather}|\textit{Cavity})\mathbf{P}(\textit{Cavity})$ în locul celor 8 ecuații

$$P(W = \textit{sunny} \wedge C = \textit{true}) = P(W = \textit{sunny}|C = \textit{true})P(C = \textit{true})$$

Full joint probability distribution: distribuția comună a tuturor variabilelor aleatoare

$$\textit{Cavity}, \textit{Toothache}, \textit{Weather} \rightarrow \mathbf{P}(\textit{Cavity}, \textit{Toothache}, \textit{Weather})$$

Introducere

Inferență

Independență

Teorema lui Bayes

Rețele bayesiene

Inferență în rețele bayesiene

Inferență probabilistă

Inferență probabilistă: calculul probabilităților condiționate, date fiind anumite observații

Exemplu: variabilele bool. *Toothache*, *Cavity*, *Catch*

Distribuția comună de probabilitate:

	toothache		\neg toothache	
	catch	\neg catch	catch	\neg catch
cavity	0.108	0.012	0.072	0.008
\neg cavity	0.016	0.064	0.144	0.576

3 variabile binare: $2^3 - 1 = 7$ parametri independenți;

pentru n variabile booleene, tabela are dimensiunea $O(2^n)$

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

Probabilitate marginală

Distribuția peste o submulțime de var. (o var.)

Marginalizare: însumăm probabilitățile pentru fiecare valoare posibilă a celorlalte variabile.

$$\mathbf{P}(\mathbf{Y}) = \sum_{z \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, z)$$

Exemplu: $\mathbf{P}(\text{Cavity}) = \sum_{z \in \{\text{Catch}, \text{Toothache}\}} \mathbf{P}(\text{Cavity}, z)$

Probabilitate marginală

$$P(\text{cavity}) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

	toothache		\neg toothache	
	catch	\neg catch	catch	\neg catch
cavity	0.108	0.012	0.072	0.008
\neg cavity	0.016	0.064	0.144	0.576

Condiționare $\mathbf{P}(\mathbf{Y}) = \sum_z \mathbf{P}(\mathbf{Y}|z)P(z)$

Probabilități condiționate: exemplu

$$\begin{aligned} P(\text{cavity}|\text{toothache}) &= \frac{P(\text{cavity} \wedge \text{toothache})}{P(\text{toothache})} \\ &= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 \end{aligned}$$

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

Obs: termenul $1/P(\text{toothache})$ const. - const. de **normalizare** pentru distribuția $\mathbf{P}(\text{Cavity}|\text{toothache})$

Probabilități condiționate: exemplu

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

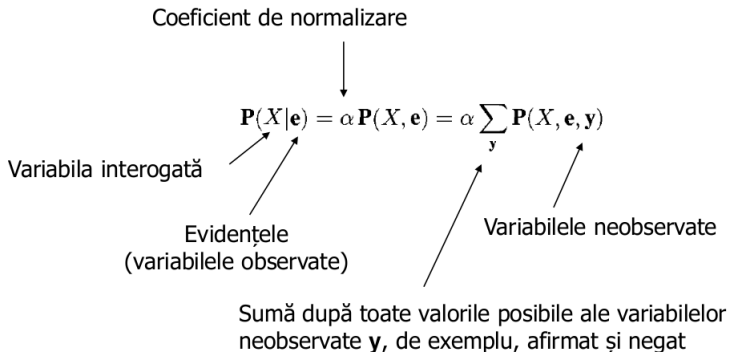
	toothache		\neg toothache	
	catch	\neg catch	catch	\neg catch
cavity	0.108	0.012	0.072	0.008
\neg cavity	0.016	0.064	0.144	0.576

Obs: putem calcula $\mathbf{P}(Cavity|toothache)$ fără a cunoaște $P(toothache)$.

Inferență

Procedura de inferență: fie variabila X (*Cavity*), \mathbf{E} lista de variabile evidență (*Toothache*), \mathbf{e} lista de valori observate, \mathbf{Y} variabile neobservate (*Catch*)

$$\mathbf{P}(X|\mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y}) \quad (1)$$



Introducere

Inferență

Independență

Teorema lui Bayes

Rețele bayesiene

Inferență în rețele bayesiene

Independență

Exemplu: Adăugăm o variabilă *Weather*.

- ▶ Regula produs

$$P(\text{toothache}, \text{catch}, \text{cavity}, \text{cloudy}) = \\ P(\text{cloudy} | \text{toothache}, \text{catch}, \text{cavity}) P(\text{toothache}, \text{catch}, \text{cavity})$$

- ▶ Obs: problemele dentare nu influențează vremea și invers

$$P(\text{cloudy} | \text{toothache}, \text{catch}, \text{cavity}) = P(\text{cloudy})$$

independență (marginală, absolută)

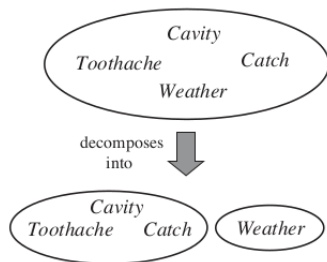
- ▶ Deducem

$$P(\text{toothache}, \text{catch}, \text{cavity}, \text{cloudy}) = \\ P(\text{cloudy}) P(\text{toothache}, \text{catch}, \text{cavity})$$

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

Independență

Variabilele X și Y sunt **independente**: $\mathbf{P}(X|Y) = \mathbf{P}(X)$ sau $\mathbf{P}(Y|X) = \mathbf{P}(Y)$ sau $\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$.



Mai puține informații necesare pentru a specifica distribuția comună de probabilitate. Distribuția comună poate fi *factorizată* în două distribuții.

Introducere

Inferență

Independență

Teorema lui Bayes

Rețele bayesiene

Inferență în rețele bayesiene

Teorema lui Bayes, '73

Utilizând regula produs:

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

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

→

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)}$$

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

Teorema lui Bayes

$$P(I|E) = \frac{P(E|I)P(I)}{P(E)}$$

- ▶ I ipoteza, E evidența (provine din datele observate)
- ▶ $P(I)$ probabilitatea a priori (*prior*) (gradul de încredere în ipoteză)
- ▶ $P(E|I)$ verosimilitatea (*likelihood*) (măsura în care s-a observat evidența, în condițiile îndeplinirii ipotezei)
- ▶ $P(I|E)$ probabilitatea a posteriori a ipotezei, dată fiind evidența (*posterior*)

Teorema lui Bayes

Cunoaștem evidența (efectul unei cauze necunoscute), și dorim să determinăm cauza:

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

Diagnostic medical: doctorul cunoaște $P(\text{symptoms}|\text{disease})$ și identifică diagnosticul $P(\text{disease}|\text{symptoms})$.

Exemplu: $P(s|m) = 0.7$ meningita cauzează înțepenirea gâtului în 70% din cazuri, $P(m) = 1/50000$, $P(s) = 0.01$.

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.7 \cdot 1/50000}{0.01} = 0.0014$$

Teorema lui Bayes

Forma generală

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

α const de normalizare

Teorema lui Bayes

Dacă avem mai mult de o variabilă evidență?

- ▶ Dacă cunoaștem distribuția comună de probabilitate
 $P(\text{Cavity}|\text{toothache} \wedge \text{catch}) = \alpha \langle 0.108, 0.016 \rangle \approx \langle 0.871, 0.129 \rangle.$

- ▶ Utilizând teorema lui Bayes:

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

- ▶ *Toothache* și *Catch* nu sunt independente
(sunt independente, dată prezența/absența cariei)
 $\mathbf{P}(\text{toothache} \wedge \text{catch}|\text{Cavity}) = \mathbf{P}(\text{toothache}|\text{Cavity}) \mathbf{P}(\text{catch}|\text{Cavity})$
Independență condiționată a *toothache* și *catch*, dat *Cavity*

Teorema lui Bayes

Independență condiționată a două variabile X și Y , dat Z :

$$\mathbf{P}(X, Y|Z) = \mathbf{P}(X|Z)\mathbf{P}(Y|Z).$$

Exemplu:

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

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

Pentru n simptome indep cond, dat $Cavity$, dimensiunea reprezentării crește liniar.

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

(modelul *Naive Bayes*)

Independență și dependență condiționată

Exemplul 1. Ion și Maria dau cu banul de 100 de ori. Fiecare are un ban diferit.

- ▶ **evenimente independente**
- ▶ rezultatul unui experiment nu influențează rezultatul celui alt experiment

Independență și independență condiționată

Exemplul 2a. Ion și Maria dau cu același ban

- ▶ dacă banul nu este corect, evenimentul A (Ion) poate aduce informații asupra evenimentului B (Maria)
- ▶ evenimentele **nu sunt independente**
(rezultatul unui experiment poate influența cunoștințele despre rezultatul celuilalt)

Independență și independență condiționată

Exemplul 2b. Fie C variabila „banul este influențat în favoarea pajurei”

- ▶ dacă știm C, experimentul A nu mai aduce informații noi asupra lui B
 $P(B|A, C) = P(B|C)$
- ▶ A și B sunt **independente condiționat**, dat fiind C
- ▶ situație numită „cauză comună”

Independență și independență condiționată

Exemplul 3. Ion și Maria locuiesc în zone diferite ale orașului și vin la serviciu cu tramvaiul, respectiv mașina

- ▶ „Ion a întârziat” și „Maria a întârziat” pot fi considerate independente
- ▶ dacă vatmanii sunt în grevă, atunci și traficul rutier crește; evenimentele sunt **independente condiționat**

Introducere

Inferență

Independență

Teorema lui Bayes

Rețele bayesiene

Inferență în rețele bayesiene

- ▶ Modele grafice probabilistice
- ▶ Reprezentarea informațiilor legate de evenimente probabilistice
 - ▶ → ne va ajuta să realizăm eficient raționamente
- ▶ Reprezintă dependențele între variabile

Rețea bayesiene

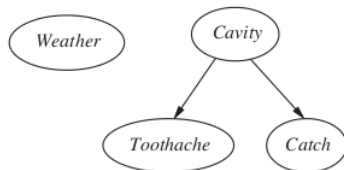
Este un digraf aciclic

- ▶ fiecare nod corespunde unei variabile aleatoare (eveniment)
- ▶ un arc de la X la Y : X este părintele lui Y (relație)
- ▶ fiecare nod X_i are o distribuție de probabilitate condiționată $\mathbf{P}(X_i | \text{Parents}(X_i))$ (efectul părinților asupra nodului)

Topologia rețelei specifică relațiile de independență condiționată

- ▶ $X \rightarrow Y$: X are o influență directă asupra lui Y .

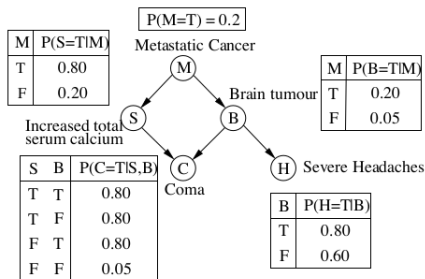
Exemplu:



Toothache și *Catch* sunt independente condiționat, dat *Cavity*

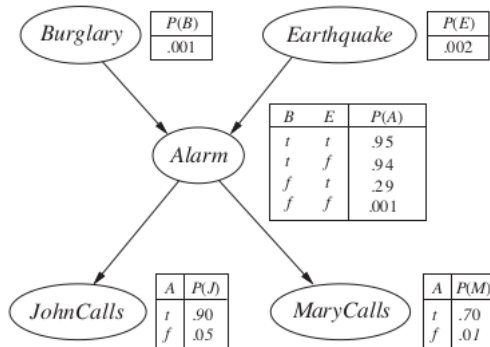
Rețele bayesiene: exemplu

Cancerul metastatic este o cauză posibilă a tumorilor cerebrale și este de asemenea, o explicație pentru creșterea calciului seric total. Oricare dintre acestea ar putea explica intrarea unui pacient în comă. Cefaleea severă este, de asemenea, asociată cu tumorile cerebrale.



Rețele bayesiene: exemplu

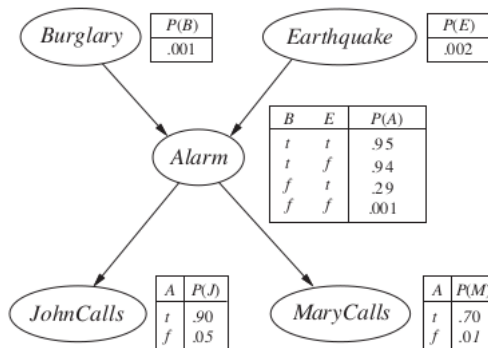
Un sistem de alarmă care sună în cazul unei spargerii, dar și în cazul unui cutremur. Vecinii John și Mary îl sună pe proprietar la serviciu dacă aud alarma.



Efracțiile si cutremurele afectează probabilitatea declanșării alarmei.
Alarma influențează probabilitatea ca John și Mary să sune.

Rețele bayesiene: exemplu

Dorim să estimăm probabilitatea unei spargerii, în funcție de cine a sunat.



Alaturat avem **tabelele de probabilitate condiționată**.

10 parametri independenți vs. 31

Interogări simple

O conjuncție $P(X_1 = x_1 \wedge \dots \wedge X_n = x_n)$

Presupunerea modelului bazat pe rețele bayesiene este că o variabilă nu depinde decât de părinții săi:

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i)) \quad (2)$$

Exemplu: probabilitatea declanșării alarmei, când nu a fost o spargere sau un cutremur, iar John și Mary au sunat

$$\begin{aligned} P(j, m, a, \neg b, \neg e) &= P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e) \\ &= 0.90 \cdot 0.70 \cdot 0.001 \cdot 0.999 \cdot 0.998 = 0.000628 \end{aligned}$$

Chain rule

Utilizăm regula produs pentru a rescrie distribuția comună de probabilitate:

$$\begin{aligned} P(x_1, \dots, x_n) &= P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1}, \dots, x_1) \\ &= P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1} | x_{n-2}, \dots, x_1) \dots P(x_2 | x_1) P(x_1) \\ &= \prod_{i=1}^n P(x_i | x_{i-1}, \dots, x_1) \end{aligned}$$

(Regula de înmulțire a probabilităților (chain rule))

E echivalentă cu

$$\mathbf{P}(X_i | X_{i-1}, \dots, X_1) = \mathbf{P}(X_i | \text{Parents}(X_i)) \quad (3)$$

$\forall X_i$ variabilă din rețea, cu condiția $\text{Parents}(X_i) \subseteq \{X_{i-1}, \dots, X_1\}$.

Rețeaua bayesiană este o reprezentare corectă dacă fiecare nod este independent condiționat de predecesorii din ordonare, dați fiind părinții.

Construcția unei rețele bayesiene

- ▶ Determină mulțimea de variabile $\{X_1, \dots, X_n\}$. Ordonează variabilele a.i. cauzele preced efectele.
- ▶ Pentru $i = 1, \dots, n$
 - ▶ alege din X_1, \dots, X_{i-1} o mulțime minimală de părinți a.i. ecuația (3) este satisfăcută
 - ▶ pentru fiecare părinte, inserează un arc de la acesta la X_i
 - ▶ adaugă tabela de probabilitate condiționată $\mathbf{P}(X_i | \text{Parents}(X_i))$

Exemplu: deoarece

$P(\text{MaryCalls} | \text{JohnCalls}, \text{Alarm}, \text{Earthquake}, \text{Burglary}) =$

$P(\text{MaryCalls} | \text{Alarm})$, rezultă *Alarm* este singurul părinte al *MaryCalls*

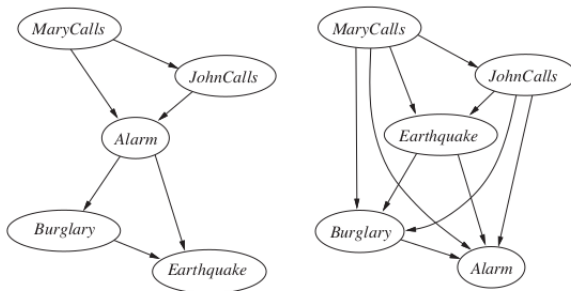
Rețele bayesiene

- Pentru n variabile, fiecare influențată de cel mult k variabile $\rightarrow 2^k$ (pentru a specifica o tabelă de probabilitate condiționată) $\rightarrow n2^k$.
vs.

Distribuția comună: 2^n

Exemplu: $n=30$ noduri, $k = 5$ părinți $\rightarrow 960$ vs. 10^9 .

- Structura rețelei depinde de ordonarea nodurilor



Sortarea topologică

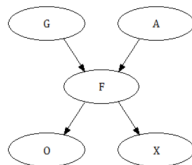
Sortarea topologică a unui graf este o ordonare liniară a nodurilor a.i. pentru fiecare arc $A \rightarrow B$, A apare înaintea lui B

- ▶ pentru o rețea bayesiană, sortarea topologică asigură faptul că părinții vor apărea înaintea fiilor
- ▶ dacă graful este orientat aciclic, există cel puțin o soluție; dacă există cicluri, sortarea topologică nu este posibilă

Sortare topologica - algoritmul lui Kahn

```
 $L \leftarrow$  Empty list that will contain the sorted elements  
 $S \leftarrow$  Set of all nodes with no incoming edge
```

```
while  $S$  is not empty do  
  remove a node  $n$  from  $S$   
  add  $n$  to  $L$   
  for each node  $m$  with an edge  $e$  from  $n$  to  $m$  do  
    remove edge  $e$  from the graph  
    if  $m$  has no other incoming edges then  
      insert  $m$  into  $S$   
  
if graph has edges then  
  return error (graph has at least one cycle)  
else  
  return  $L$  (a topologically sorted order)
```

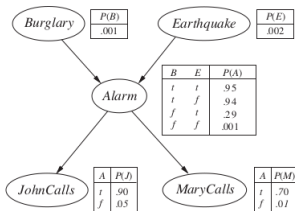


1. $L = \emptyset, S = \{G, A\}$
2. $L = \{G\}, S = \{A\}$
3. se elimină arcul GF
F nu poate fi adăugat în S ($\exists AF$)
4. $L = \{G, A\}, S = \emptyset$
5. se elimină AF, $S = \{F\}$
6. $L = \{G, A, F\}, S = \emptyset$
7. se elimină FO, $S = \{O\}, \dots$
 $\rightarrow L = \{G, A, F, O, X\}$

Complexitate timp: $O(n + m)$, n noduri, m arce

Independență condiționată

- Fiecare variabilă este independentă condiționat de ne-descendenți, dați părinții.



JohnCalls este independent de *Burglary*, *Earthquake*, *MarryCalls*, dat *Alarm*.

- **Markov blanket**: un nod e independent conditionat de celălalte noduri, dați părinții, copiii și părinții copiilor

Burglary este independent de *JohnCalls* și *MaryCalls*, dat *Alarm* și *Earthquake*

Introducere

Inferență

Independență

Teorema lui Bayes

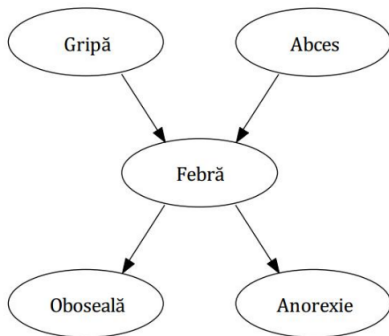
Rețele bayesiene

Inferență în rețele bayesiene

Inferența probabilităților marginale

- ▶ Calculează probabilitățile nodurilor, în lipsa unor noduri evidență
- ▶ Pentru un nod, calculăm suma probabilităților condiționate de combinațiile posibile de valori ale părinților, înmulțite cu probabilitățile părinților de a avea valorile respective.

Inferența probabilităților marginale: exemplu



$P(\text{Gripă} = \text{Da})$	$P(\text{Gripă} = \text{Nu})$
0,1	0,9

$P(\text{Abces} = \text{Da})$	$P(\text{Abces} = \text{Nu})$
0,05	0,95

Gripă	Abces	$P(\text{Febră} = \text{Da})$	$P(\text{Febră} = \text{Nu})$
Da	Da	0,8	0,2
Da	Nu	0,7	0,3
Nu	Da	0,25	0,75
Nu	Nu	0,05	0,95

Febră	$P(\text{Oboseală} = \text{Da})$	$P(\text{Oboseală} = \text{Nu})$
Da	0,6	0,4
Nu	0,2	0,8

Febră	$P(\text{Anorexie} = \text{Da})$	$P(\text{Anorexie} = \text{Nu})$
Da	0,5	0,5
Nu	0,1	0,9

Nodul Febră

$$P(F_D) =$$

$$P(F_D|G_D, A_D) \cdot P(G_D) \cdot P(A_D) +$$

$$P(F_D|G_D, A_N) \cdot P(G_D) \cdot P(A_N) +$$

$$P(F_D|G_N, A_D) \cdot P(G_N) \cdot P(A_D) +$$

$$P(F_D|G_N, A_N) \cdot P(G_N) \cdot P(A_N) =$$

$$0,8 \cdot 0,1 \cdot 0,05 + 0,7 \cdot 0,1 \cdot 0,95 + 0,25 \cdot 0,9 \cdot 0,05 + 0,05$$

$$\cdot 0,9 \cdot 0,95 = 0,1245 \approx 12\%.$$

$$P(F_N) = 1 - P(F_D) = 0,8755 \approx 88\%.$$

$$F_D = (\text{Febră} = \text{Da})$$

$$F_N = (\text{Febră} = \text{Nu})$$

Într-o rețea bayesiană,
un nod poate avea
oricâte valori posibile,
nu doar două

$$P(O_D) =$$

$$P(O_D|F_D) \cdot P(F_D) + P(O_D|F_N) \cdot P(F_N) =$$

$$0,6 \cdot 0,1245 + 0,2 \cdot 0,8755 = 0,2498 \approx 25\%,$$

$$P(O_N) = 1 - P(O_D) = 0,7502 \approx 75\%.$$

$O_D = (\text{Oboselă} = \text{Da})$

$O_N = (\text{Oboselă} = \text{Nu})$

$$P(X_D) =$$

$$P(X_D|F_D) \cdot P(F_D) + P(X_D|F_N) \cdot P(F_N) =$$

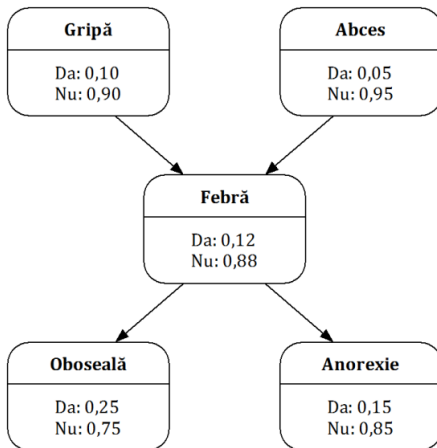
$$0,5 \cdot 0,1245 + 0,1 \cdot 0,8755 = 0,1498 \approx 15\%,$$

$$P(X_N) = 1 - P(X_D) = 0,8502 \approx 85\%.$$

$X_D = (\text{Anorexie} = \text{Da})$

$X_N = (\text{Anorexie} = \text{Nu})$

Probabilitățile marginale ale nodurilor



Inferență în rețele bayesiene

- ▶ Sistem de inferență probabilist: calculăm distribuția de probabilitate aposteriori pentru o mulțime de variabile, dat un eveniment
- ▶ X variabila, \mathbf{E} mulțimea de variabile evidență E_1, \dots, E_m , \mathbf{e} eveniment, \mathbf{Y} non-evidență, Y_1, \dots, Y_l variabile ascunse

Distribuția de probabilitate aposteriori $\mathbf{P}(X|\mathbf{e}) = ?$

Exemplu: observăm evenimentul în care $JohnCalls = true$ și $MaryCalls = true$; atunci probabilitatea unei efracții:

$\mathbf{P}(Burglary|JohnCalls = true, MaryCalls = true) = \langle 0.284, 0.716 \rangle$.

Probabilitatea condiționată

$$\mathbf{P}(X|\mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y}) \quad (1)$$

O rețea bayesiană oferă o reprezentare a distribuției comune.

Conform ecuației (2), termenul $\mathbf{P}(X, \mathbf{e}, \mathbf{y})$ poate fi scris ca produs de probabilități condiționate.

Inferență prin enumerare

Exemplu: $P(\text{Burglary} | \text{JohnCalls} = \text{true}, \text{MaryCalls} = \text{true})$

- ▶ Variabilele ascunse sunt *Earthquake* și *Alarm*.

Conform ecuației (3),

$$\mathbf{P}(B|j, m) = \alpha \mathbf{P}(B, j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, j, m, e, a)$$

- ▶ Pentru $\text{Burglary} = \text{true}$, conform ecuației (2):

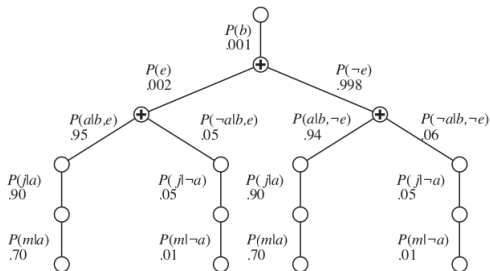
$$P(b|j, m) = \alpha \sum_e \sum_a P(b)P(e)P(a|b, e)P(j|a)P(m|a)$$

Termenul $P(b)$ const, $P(e)$ nu depinde de a :

$$P(b|j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a|b, e)P(j|a)P(m|a) = \alpha \times 0.00059224$$

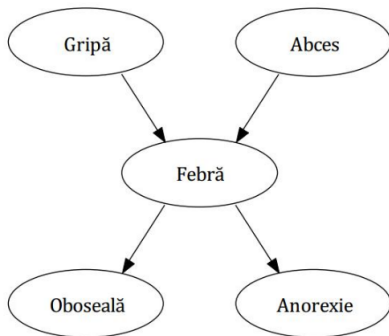
- ▶ $P(B|j, m) = \alpha \langle 0.00059224, 0.0014919 \rangle \approx \langle 0.284, 0.716 \rangle$.
Probabilitatea unei spargerii: 28%.

Procesul de evaluare:



Inferență prin enumerare: exemplu

Interogare: Care este probabilitatea ca o persoană să aibă gripă, dacă prezintă simptome de oboseală și anorexie?



$P(\text{Gripă} = \text{Da})$	$P(\text{Gripă} = \text{Nu})$
0,1	0,9

$P(\text{Abces} = \text{Da})$	$P(\text{Abces} = \text{Nu})$
0,05	0,95

Gripă	Abces	$P(\text{Febră} = \text{Da})$	$P(\text{Febră} = \text{Nu})$
Da	Da	0,8	0,2
Da	Nu	0,7	0,3
Nu	Da	0,25	0,75
Nu	Nu	0,05	0,95

Febră	$P(\text{Oboseală} = \text{Da})$	$P(\text{Oboseală} = \text{Nu})$
Da	0,6	0,4
Nu	0,2	0,8

Febră	$P(\text{Anorexie} = \text{Da})$	$P(\text{Anorexie} = \text{Nu})$
Da	0,5	0,5
Nu	0,1	0,9

Inferență prin enumerare: exemplu

- ▶ Calculăm independent $P(G_D|O_D, X_D)$ și $P(G_N|O_D, X_D)$
- ▶ Pentru $P(G_D|O_D, X_D)$, variabilele rămase sunt *Abcesul* și *Febra*
- ▶ Sumăm probabilitățile corespunzătoare tuturor valorilor acestor variabile: $a \in \{A_D, A_N\}$ și $f \in \{F_D, F_N\}$
- ▶ Pentru a crește eficiența calculelor, se recomandă ca variabilele rămase să fie mai întâi sortate topologic, a.i. părinții să apară înaintea copiilor.
În acest caz, se vor putea descompune mai ușor sumele, scoțând în față factorii care nu depind de o anumită variabilă.

Inferență prin enumerare: exemplu

$$P(G_D|O_D, X_D) =$$

$$\alpha \cdot \sum_{a \in \{A_D, A_N\}} \sum_{f \in \{F_D, F_N\}} P(G_D, a, f, O_D, X_D) =$$

$$\alpha \cdot \sum_a \sum_f P(G_D) \cdot P(a) \cdot P(f|G_D, a) \cdot P(O_D|f) \cdot P(X_D|f) =$$

$$\alpha \cdot P(G_D) \cdot \sum_a P(a) \cdot \sum_f P(f|G_D, a) \cdot P(O_D|f) \cdot P(X_D|f) =$$

$$\alpha \cdot P(G_D) \cdot \sum_a P(a) \cdot [P(F_D|G_D, a) \cdot P(O_D|F_D) \cdot P(X_D|F_D) +$$

$$P(F_N|G_D, a) \cdot P(O_D|F_N) \cdot P(X_D|F_N)] =$$

$$\alpha \cdot P(G_D) \cdot \{P(A_D) \cdot [P(F_D|G_D, A_D) \cdot P(O_D|F_D) \cdot P(X_D|F_D) +$$

$$P(F_N|G_D, A_D) \cdot P(O_D|F_N) \cdot P(X_D|F_N)] +$$

$$P(A_N) \cdot [P(F_D|G_D, A_N) \cdot P(O_D|F_D) \cdot P(X_D|F_D) +$$

$$P(F_N|G_D, A_N) \cdot P(O_D|F_N) \cdot P(X_D|F_N)]\} =$$

$$\alpha \cdot 0,1 \cdot \{0,05 \cdot [0,8 \cdot 0,6 \cdot 0,5 + 0,2 \cdot 0,2 \cdot 0,1] +$$

$$0,95 \cdot [0,7 \cdot 0,6 \cdot 0,5 + 0,3 \cdot 0,2 \cdot 0,1]\} =$$

$$\alpha \cdot 0,02174.$$

Inferență prin enumerare: exemplu

$$\begin{aligned}P(G_N|O_D, X_D) &= \\ \alpha \cdot \sum_{a \in \{A_D, A_N\}} \sum_{f \in \{F_D, F_N\}} P(G_N, a, f, O_D, X_D) &= \\ \alpha \cdot \sum_a \sum_f P(G_N) \cdot P(a) \cdot P(f|G_N, a) \cdot P(O_D|f) \cdot P(X_D|f) &= \\ \alpha \cdot P(G_N) \cdot \sum_a P(a) \cdot \sum_f P(f|G_N, a) \cdot P(O_D|f) \cdot P(X_D|f) &= \\ \alpha \cdot P(G_N) \cdot \{P(A_D) \cdot [P(F_D|G_N, A_D) \cdot P(O_D|F_D) \cdot P(X_D|F_D)] + \\ &\quad P(F_N|G_N, A_D) \cdot P(O_D|F_N) \cdot P(X_D|F_N)] + \\ &\quad P(A_N) \cdot [P(F_D|G_N, A_N) \cdot P(O_D|F_D) \cdot P(X_D|F_D)] + \\ &\quad P(F_N|G_N, A_N) \cdot P(O_D|F_N) \cdot P(X_D|F_N)]\} = \\ \alpha \cdot 0,9 \cdot \{0,05 \cdot [0,25 \cdot 0,6 \cdot 0,5 + 0,75 \cdot 0,2 \cdot 0,1] + \\ &\quad 0,95 \cdot [0,05 \cdot 0,6 \cdot 0,5 + 0,95 \cdot 0,2 \cdot 0,1]\} = \\ \alpha \cdot 0,03312.\end{aligned}$$

Inferență prin enumerare: exemplu

- $P(G_D|O_D, X_D) + P(G_N|O_D, X_D) = 1$
- $P(G_D|O_D, X_D) = \alpha \cdot 0,02174$
- $P(G_N|O_D, X_D) = \alpha \cdot 0,03312$
- $\Rightarrow \alpha = 18,23$

- $P(G_D|O_D, X_D) = 0,39628 \approx 40\%$
- $P(G_N|O_D, X_D) = 0,60372 \approx 60\%$

Inferență prin enumerare

function ENUMERATION-ASK(X, \mathbf{e}, bn) **returns** a distribution over X

inputs: X , the query variable

\mathbf{e} , observed values for variables \mathbf{E}

bn , a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* $\mathbf{Y} = \text{hidden variables}$ */

$Q(X) \leftarrow$ a distribution over X , initially empty

for each value x_i of X **do**

$Q(x_i) \leftarrow$ ENUMERATE-ALL($bn.VARS, \mathbf{e}_{x_i}$)

where \mathbf{e}_{x_i} is \mathbf{e} extended with $X = x_i$

return NORMALIZE($Q(X)$)

function ENUMERATE-ALL($vars, \mathbf{e}$) **returns** a real number

if EMPTY?($vars$) **then return** 1.0

$Y \leftarrow$ FIRST($vars$)

if Y has value y in \mathbf{e}

then return $P(y \mid \text{parents}(Y)) \times$ ENUMERATE-ALL(REST($vars$), \mathbf{e})

else return $\sum_y P(y \mid \text{parents}(Y)) \times$ ENUMERATE-ALL(REST($vars$), \mathbf{e}_y)

where \mathbf{e}_y is \mathbf{e} extended with $Y = y$

Algoritmul evaluează arborii de expresii în manieră DFS. Complexitatea spațiu: $O(n)$, n variabile. Complexitatea timp: $O(2^n)$ pentru o rețea cu n variabile bool.

Algoritmul de eliminare a variabilelor (*variable elimination*)

- ▶ Obs: $P(j|a)P(m|a)$ și $P(j|\neg a)P(m|\neg a)$ sunt calculate de două ori, pentru fiecare valoare a lui e .
- ▶ Idee: realizează calculele de la dreapta către stânga (de jos în sus) și salvează rezultatele.

Variable elimination: exemplu

$$P(B|j, m) = \alpha \underbrace{P(b)}_{f_1(B)} \sum_e \underbrace{P(e)}_{f_2(E)} \sum_a \underbrace{P(a|B, e)}_{f_3(A, B, E)} \underbrace{P(j|a)}_{f_4(A)} \underbrace{P(m|a)}_{f_5(A)}$$

- fiecare factor f_i este o matrice indexată de variabilele argument:

$$f_4(A) = \begin{pmatrix} P(j|a) \\ P(j|\neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix}, f_5(A) = \begin{pmatrix} P(m|a) \\ P(m|\neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix}.$$

$f_3(A, B, E)$ este o matrice $2 \times 2 \times 2$.

$$P(B|j, m) = \alpha f_1(B) \times \sum_e f_2(E) \times \sum_a f_3(A, B, E) \times f_4(A) \times f_5(A)$$

Variable elimination: exemplu

$$\begin{aligned}f_6(B, E) &= \sum_a f_3(A, B, E) \times f_4(A) \times f_5(A) \\&= f_3(a, B, E) \times f_4(a) \times f_5(a) + f_3(\neg a, B, E) \times f_4(\neg a) \times f_5(\neg a)\end{aligned}$$

$$\begin{aligned}f_7(B) &= \sum_e f_2(E) \times f_6(B, E) \\&= f_2(e) \times f_6(B, e) + f_2(\neg e) \times f_6(B, \neg e)\end{aligned}$$

$$P(B|j, m) = \alpha f_1(B) \times f_7(B)$$