Inferenza Probabilistica e Reti Bayesiane

Corso di Sistemi Intelligenti,

Russell & Norvig, Capitoli 13 e 14

Probabilità condizionale

Probabilità a posteriori o condizionale

p.e., $P(\textit{Cavità}|Mal\_di\_denti) = 0.8$

cioè, dato che Mal_di_denti è tutto quello che so

Se so di più, p.e., *Cavità* è anche data, allora abbiamo

$P(\textit{Cavità}|Mal\_di\_denti, \textit{Cavità}) = 1$

Nota: la credenza meno specifica *rimane valida* dopo che nuova evidenza arriva, ma non necessariamente rimane *utile*

Nuova evidenza può essere irrilevante, permettendo semplificazioni, p.e.,

$P(\textit{Cavità}|Mal\_di\_denti, Vince\_Inter) = P(\textit{Cavità}|Mal\_di\_denti) = 0.8$

Questo tipo di inferenza, dovuta alla conoscenza del dominio, è cruciale

Probabilità condizionale

Definizione di probabilità condizionale:

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

La regola del prodotto fornisce una definizione alternativa:

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

Una versione generale vale sulle distribuzioni, p.e.,

$$\mathbf{P}(Tempo, \mathit{Cavità}) = \mathbf{P}(Tempo|\mathit{Cavità})\mathbf{P}(\mathit{Cavità})$$

(Visto come un insieme 4×2 di equazioni, *no* moltiplicazione di matrici)

Chain rule è derivata dalla applicazione ripetuta della regola del prodotto:

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

Inferenza tramite enumerazione

Si inizia con la distribuzione congiunta:

	$toothache$		$\neg\, toothache$	
	$catch$	$\neg\, catch$	$catch$	$\neg\, catch$
$cavity$	**.108**	**.012**	**.072**	**.008**
$\neg\, cavity$	**.016**	**.064**	**.144**	**.576**

Per ogni proposizione ϕ, si sommano gli eventi atomici dove essa è vera:

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

Inferenza tramite enumerazione

Si inizia con la distribuzione congiunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Per ogni proposizione ϕ, si sommano gli eventi atomici dove essa è vera:

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

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

Inferenza tramite enumerazione

Si inizia con la distribuzione congiunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Per ogni proposizione ϕ, si sommano gli eventi atomici dove essa è vera:

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

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

Inferenza tramite enumerazione

Si inizia con la distribuzione congiunta:

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

Si possono calcolare anche le probabilità condizionali:

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

Normalizzazione

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Il denominatore può essere visto come una *costante di normalizzazione* α

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

Idea generale: calcolare la distribuzione sulla variabile della query fissando le variabili di evidenza e sommando sulle variabili nascoste

Inferenza tramite enumerazione

Tipicamente siamo interessati a

la distribuzione congiunta a posteriori delle variabili di query $\mathbf{Y}$
dati specifici valori $\mathbf{e}$ per le variabili di evidenza $\mathbf{E}$

Poniamo le variabili nascoste essere $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Allora la somma desiderata di entrate congiunte è ottenuta sommando sulle variabili nascoste:

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

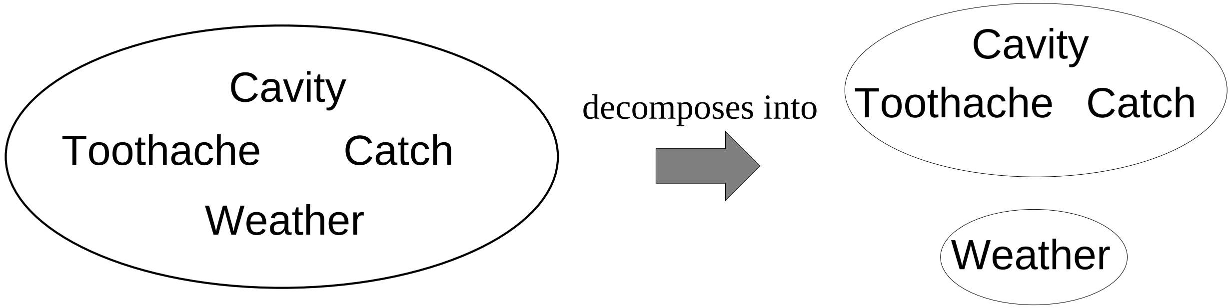
I termini nella sommatoria sono entrate congiunte perché $\mathbf{Y}$, $\mathbf{E}$, e $\mathbf{H}$ insieme esauriscono l'insieme delle variabili aleatorie

Problemi ovvi:

1) Complessità caso pessimo in tempo $O(d^n)$ dove d è l'arietà più grande
2) Complessità in spazio $O(d^n)$ per memorizzare la distribuzione congiunta
3) Come stabilire i valori per $O(d^n)$ entrate???

Indipendenza

A e B sono indipendenti sse
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



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

32 entrate ridotte a 12; per n monete "truccate" indipendenti, $2^n \to n$

Indipendenza assoluta potente ma rara

Nei problemi reali sono coinvolte centinaia di variabili,
nessuana delle quali è indipendente. Che fare?

Indipendenza condizionale

$\mathbf{P}(Toothache, Cavity, Catch)$ ha $2^3 - 1 = 7$ entrate indipendenti

Se si ha una cavità, la probabilità che la sonda si fermi su in essa non dipende dal fatto di avere il mal di denti:

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

La stessa indipendenza vale se non c'è la cavità:

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

$Catch$ è *condizionalmente indipendente* da $Toothache$ dato $Cavity$:

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

Affermazione equivalente:

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

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

Indipendenza condizionale

Scrivere la distribuzione congiunta completa usando la chain rule:

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

Cioè, $2 + 2 + 1 = 5$ numeri indipendenti

In molti casi, l'uso di indipendenza condizionale riduce la dimensione della rappresentazione della probabilità congiunta da essere esponenziale in n a lineare n.

L'indipendenza condizionale è la forma più basilare e robusta di conoscenza sugli ambienti incerti.

Regola di Bayes

Regola del prodotto $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$\Rightarrow$ Bayes' rule $P(a|b) = \dfrac{P(b|a)P(a)}{P(b)}$

o in forma di distribuzione

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

Utile per ottenere probabilità diagnostica a partire da probabilità causale:

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

P.e., sia M la rappresentazione di Meningite, e S di collo rigido:

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

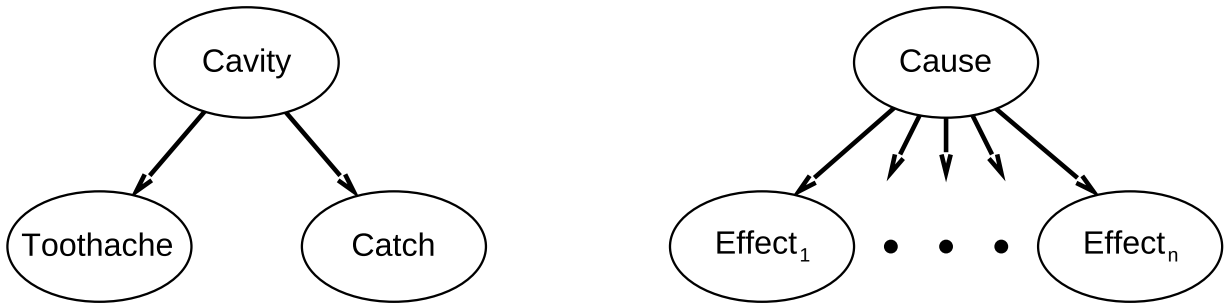
Nota: la probabilità a posteriori della Meningite ancora piccola!

Regola di Bayes e indipendenza condizionale

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

Questo è un esempio di modello *naive Bayes*:

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



Il numero totale di parametri è *lineare* in n

Reti Bayesiane (Bayesian networks)

Una semplice notazione grafica per asserzioni condizionalmente indipendenti e quindi per specifiche di distribuzioni condizionali complete Sintassi:

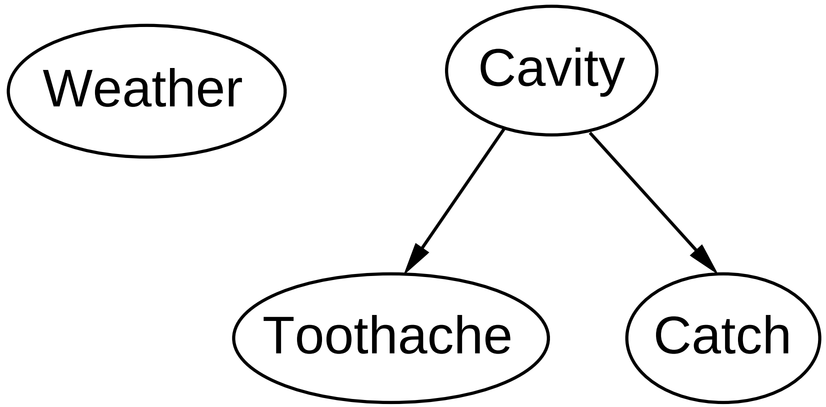
- un insieme di nodi, uno per variabile
- un grafo diretto aciclico (link $\approx$ "influenza direttamente")
- una distribuzione condizionale per ogni nodo dati i suoi genitori:

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

Nel caso più semplice, distribuzione condizionale rappresentata come una tabella della probabilità condizionale (CPT) data la distribuzione su X_i per ogni combinazione di valori assunti dai genitori

Esempio

La topologia della rete codifica asserzioni di indipendenza condizionale:



$Weather$ è indipendente dalle altre variabili

$Toothache$ e $Catch$ sono condizionalmente indipendenti data $Cavity$

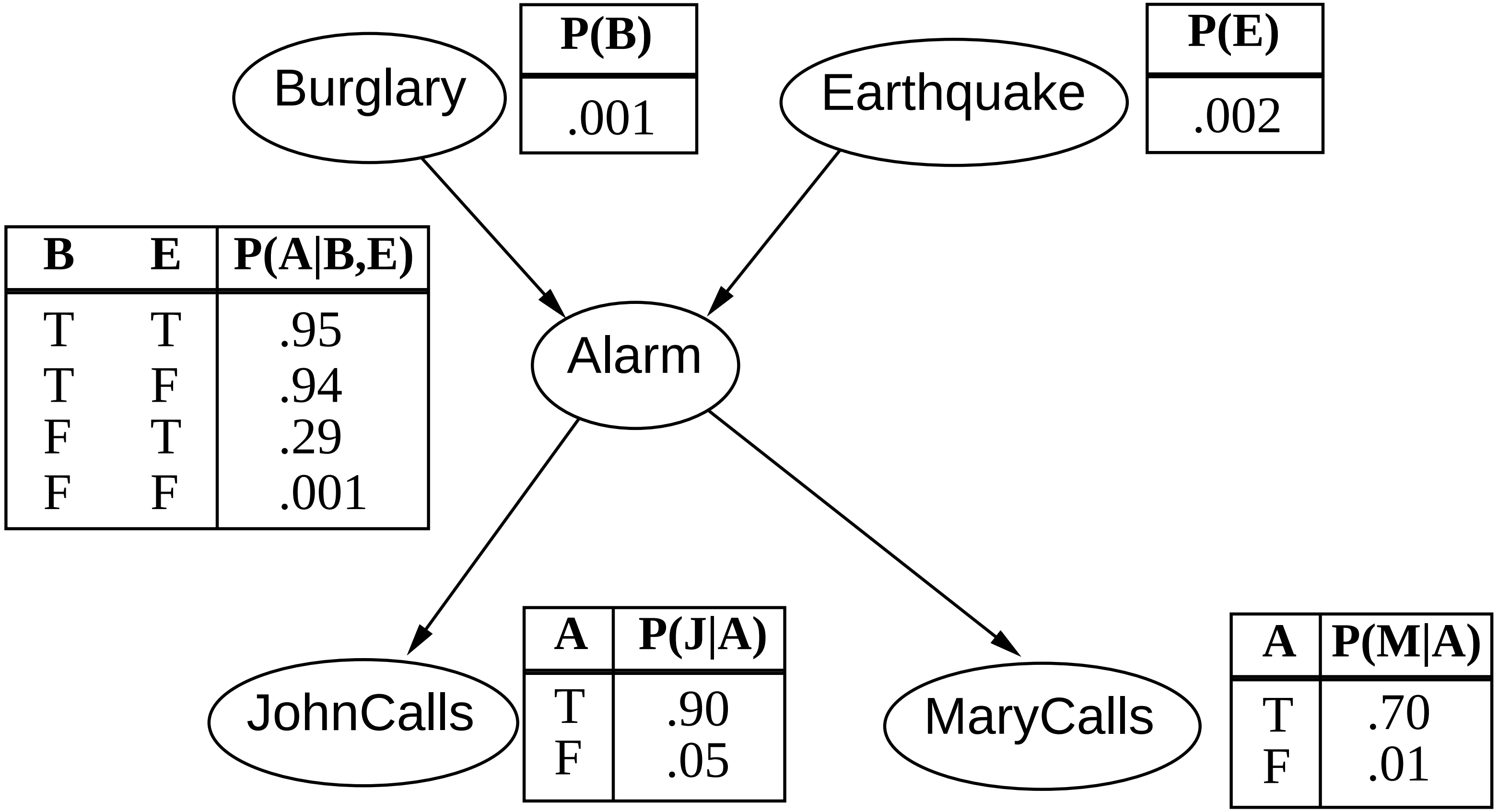
Esempio

Sono al lavoro, il vicino John chiama per dire che il mio allarme $Alarm$ è entrato in funzione, ma la vicina Mary non chiama. Alcune volte l'allarme è attivato da piccole scosse di terremoto. C'è un ladro in casa ?

Variabili: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
La topologia della rete riflette conoscenza "causale":

- Un ladro può attivare l'allarme
- Un terremoto può attivare l'allarme
- L'attivazione dell'allarme può indurre Mary a chiamare
- L'attivazione dell'allarme può indurre John a chiamare

Esempio



P(B)
.001

P(E)
.002

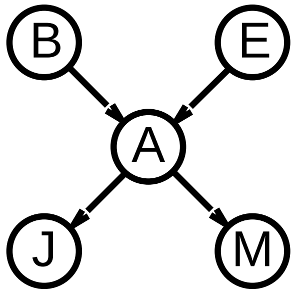
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

Compattezza

Una CPT per variabili Booleane X_i con k genitori Booleani ha 2^k righe per le combinazioni di valori dei genitori



Ogni riga richiede un numero p per $X_i = vero$
(il numero per $X_i = falso$ è $1 - p$)

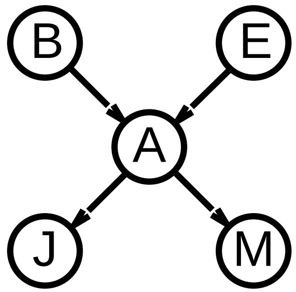
Se ogni variabile non ha più di k genitori,
la rete completa richiede $O(n \cdot 2^k)$ numeri

Cioè, cresce linearmente con n, vs. $O(2^n)$ per la distribuzione congiunta completa

Per la rete precedente, $1 + 1 + 4 + 2 + 2 = 10$ numeri (vs. $2^5 - 1 = 31$)

Semantica globale

La semantica globale definisce la distribuzione congiunta completa come il prodotto delle distribuzioni condizionali locali:



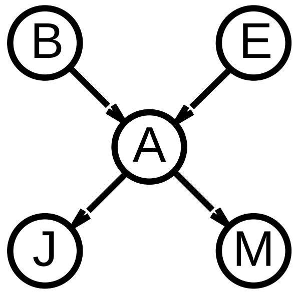
$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

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

$$=$$

Semantica globale

La semantica globale definisce la distribuzione congiunta completa come il prodotto delle distribuzioni condizionali locali:



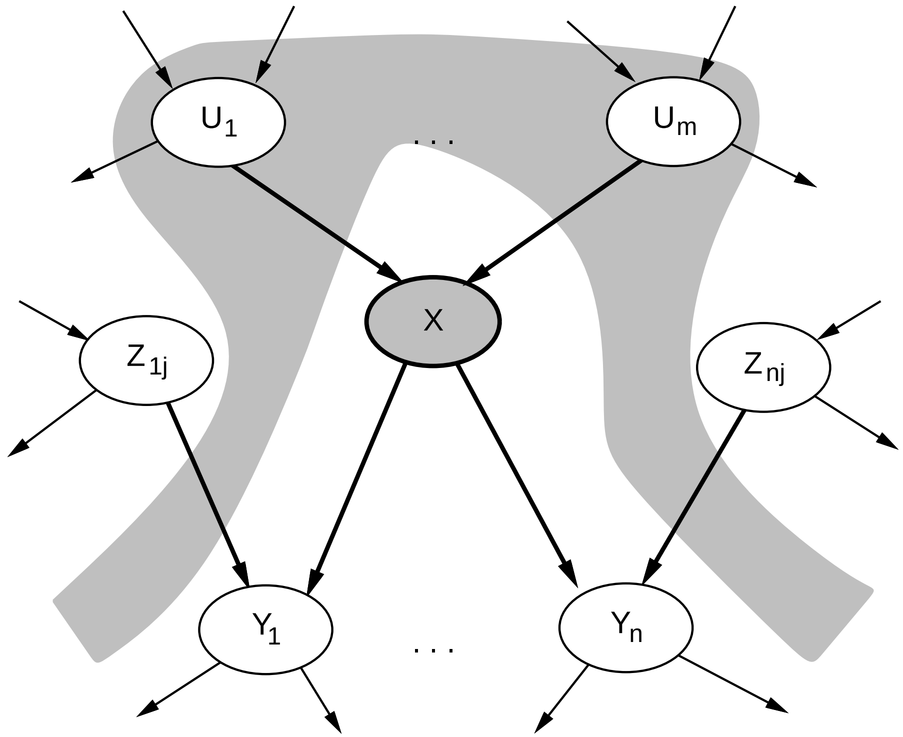
$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

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

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

Semantica locale

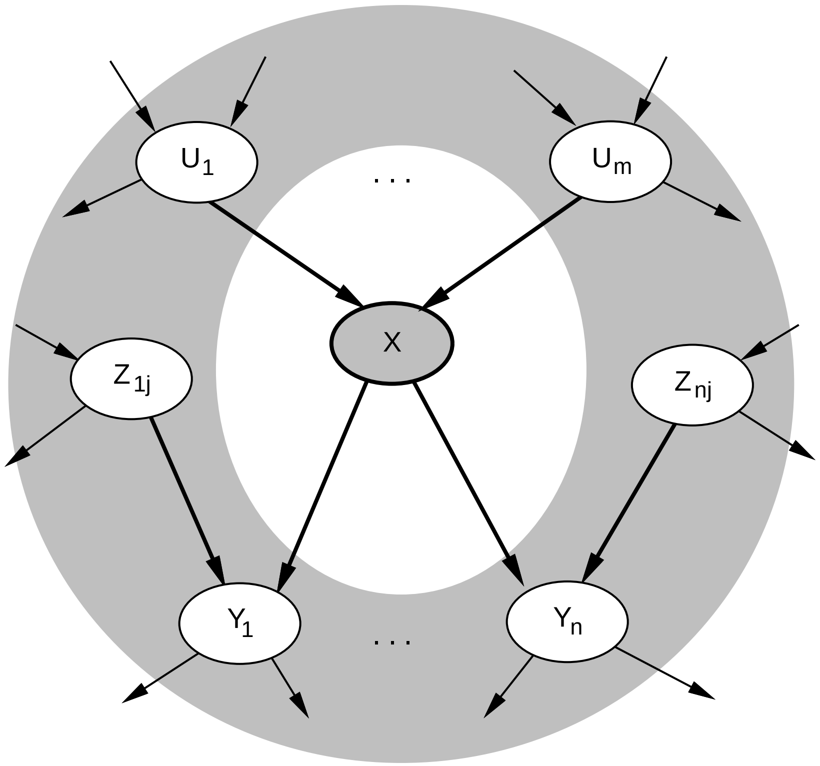
Semantica locale: ogni nodo è condizionalmente indipendente dai suoi non discendenti dati i genitori



Teorema: Semantica locale $\Leftrightarrow$ semantica globale

Markov blanket

Ogni nodo è condizionalmente indipendente da tutti gli altri dato il suo Markov blanket: genitori + figli + genitori dei figli



Construzione di Reti Bayesiane

Necessità di un metodo tale che data una serie di asserzioni di indipendenza condizionale localmente controllabili, garantisca la semantica globale desiderata

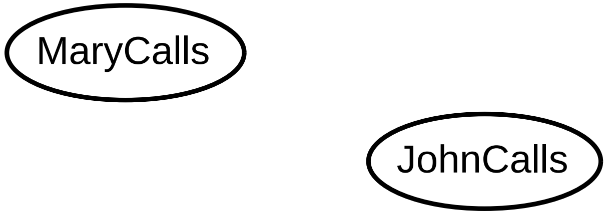
1. Scegliere un ordinamento di variabili $X_1, \ldots, X_n$
2. For $i = 1$ to n
   aggiungi X_i alla rete
   seleziona genitori da $X_1, \ldots, X_{i-1}$ tali che
   $\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_1, \ldots, X_{i-1})$

Questa scelta di genitori garantisce la semantica globale:

$$\begin{aligned} \mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1}) \quad \text{(chain rule)} \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i)) \quad \text{(per costruzione)} \end{aligned}$$

Esempio

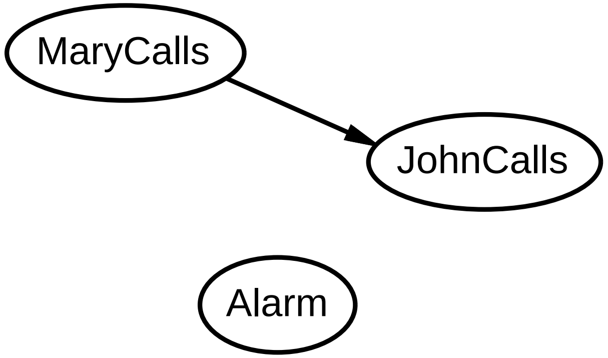
Supponiamo di scegliere l'ordine M, J, A, B, E



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

Esempio

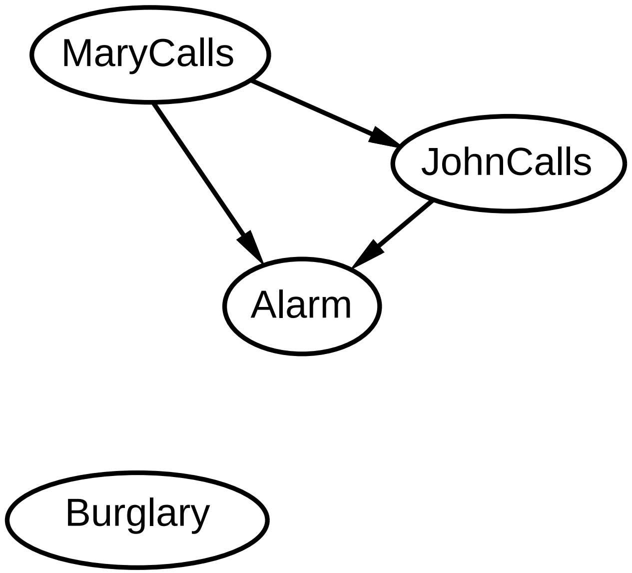
Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$?

Esempio

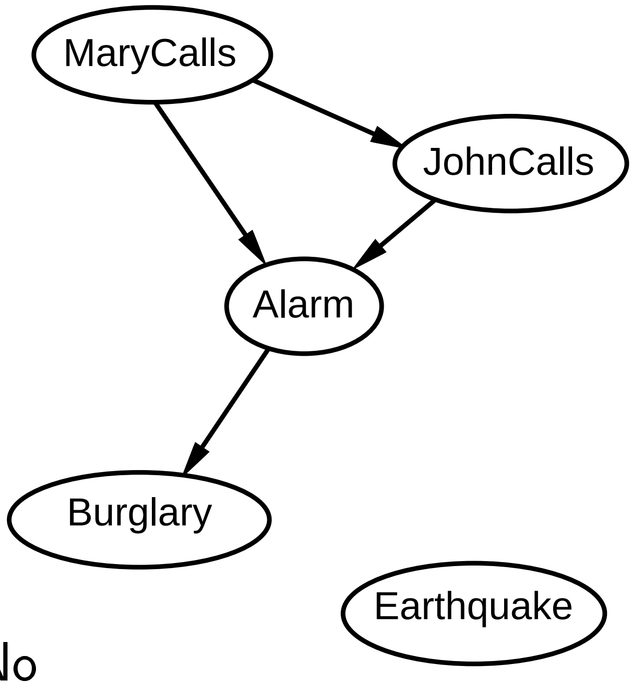
Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No
$P(B|A,J,M) = P(B|A)$?
$P(B|A,J,M) = P(B)$?

Esempio

Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$? No

$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No

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

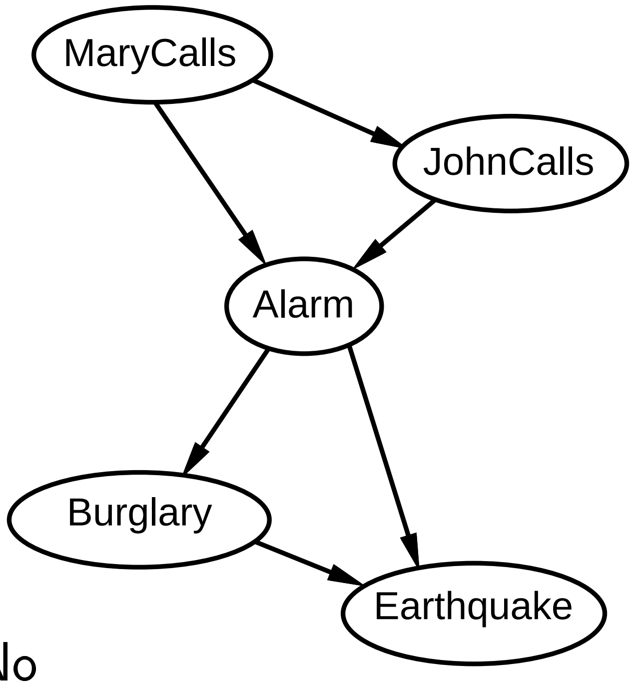
$P(B|A,J,M) = P(B)$? No

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

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

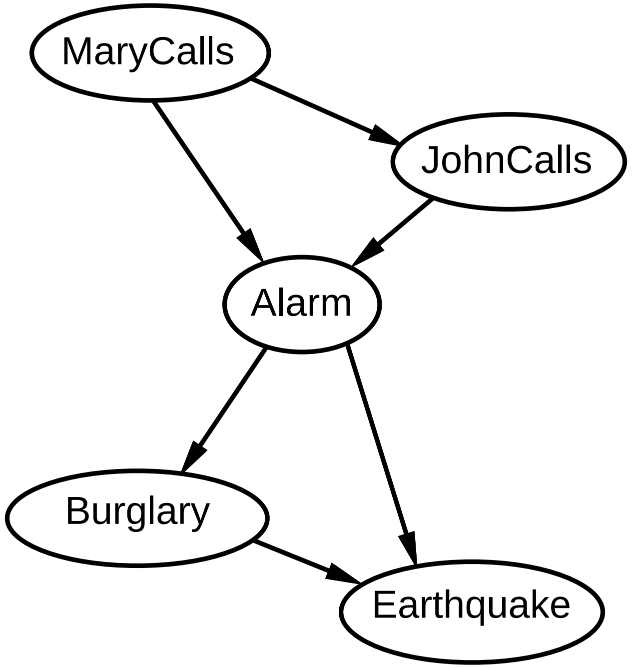
Esempio

Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No
$P(B|A,J,M) = P(B|A)$? Yes
$P(B|A,J,M) = P(B)$? No
$P(E|B,A,J,M) = P(E|A)$? No
$P(E|B,A,J,M) = P(E|A,B)$? Yes

Esempio



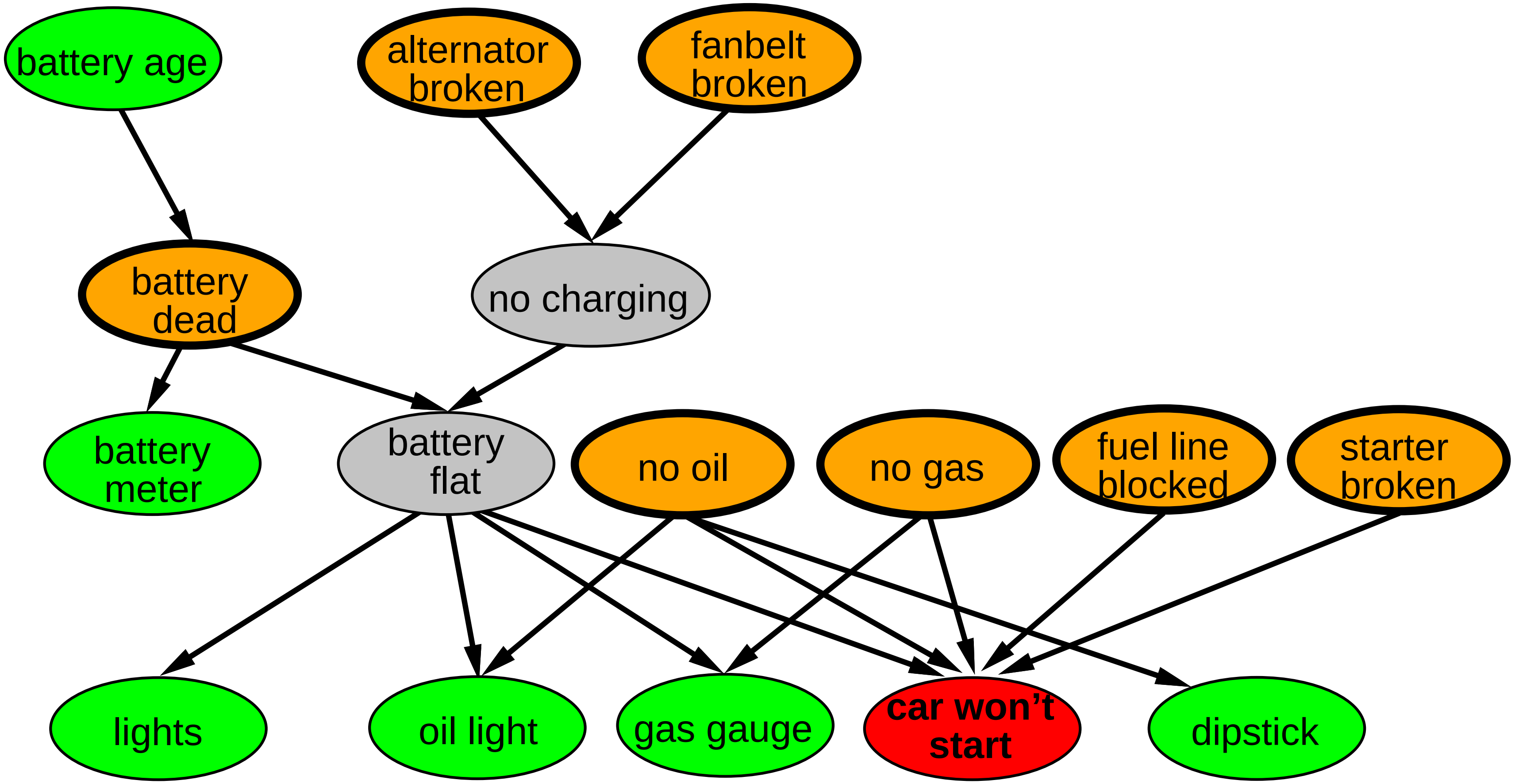
Decidere l'indipendenza condizionale è difficile nelle direzioni non causali

Valutare le probabilità condizionali è difficile in direzioni non causali

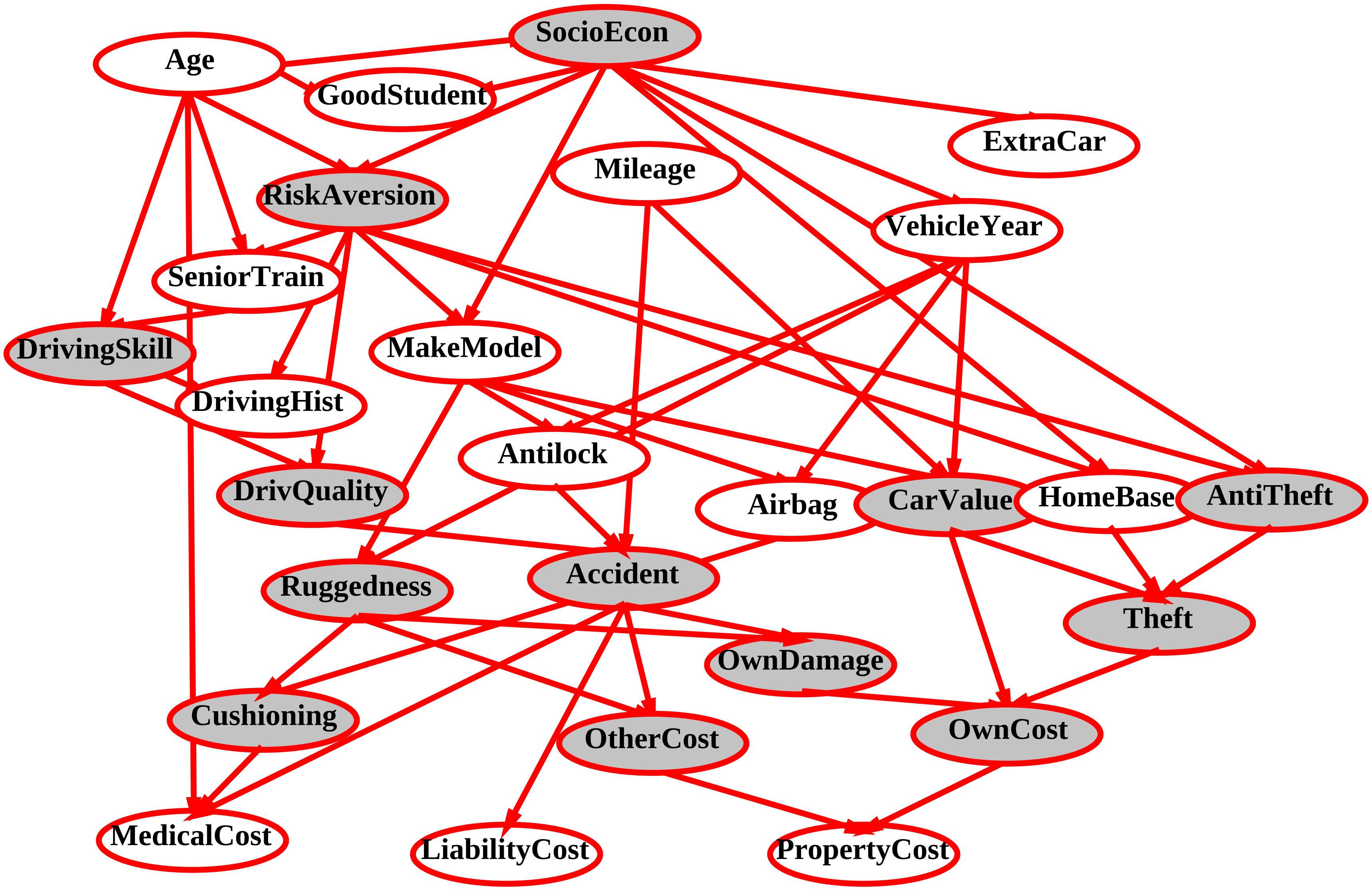
La rete è meno compatta: $1 + 2 + 4 + 2 + 4 = 13$ numeri necessari

Esempio: diagnosi per automobile

Evidenza iniziale: auto non parte
Variabili controllabili (in verde), variabili “rotto, da aggiustare” (in arancio)
Variabili nascoste (in grigio) assicurano struttura sparsa, riducono i parametri



Esempio: assicurazione dell'automobile



Compiti di inferenza

Query semplici: calcolare la probabilità a posteriori marginale $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$
p.e., $P(NoGas|Gauge=empty, Lights=on, Starts=false)$

Query congiuntive: $\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$

Decisioni ottimali: reti di decisioni includono informazioni di utilità;
inferenza probabilistica richiesta per $P(outcome|action, evidence)$

Recupero informazione: quale evidenza si deve cercare?

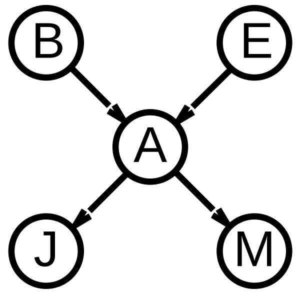
Analisi della sensitività: quali valori di probabilità sono i più critici?

Spiegazione: perché ho bisogno di un nuovo motore di avviamento?

Inferenza tramite enumerazione

Modo un pò più furbo per marginalizzare alcune variabili dalla distribuzione congiunta senza costruire esplicitamente la sua rappresentazione

Query semplice sulla rete dell'allarme:



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

Riscrittura di entrate della distribuzione congiunta usando il prodotto di entrate di CPT:

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

Enumerazione ricorsiva depth-first: $O(n)$ in spazio, $O(d^n)$ in tempo

Algoritmo di enumerazione

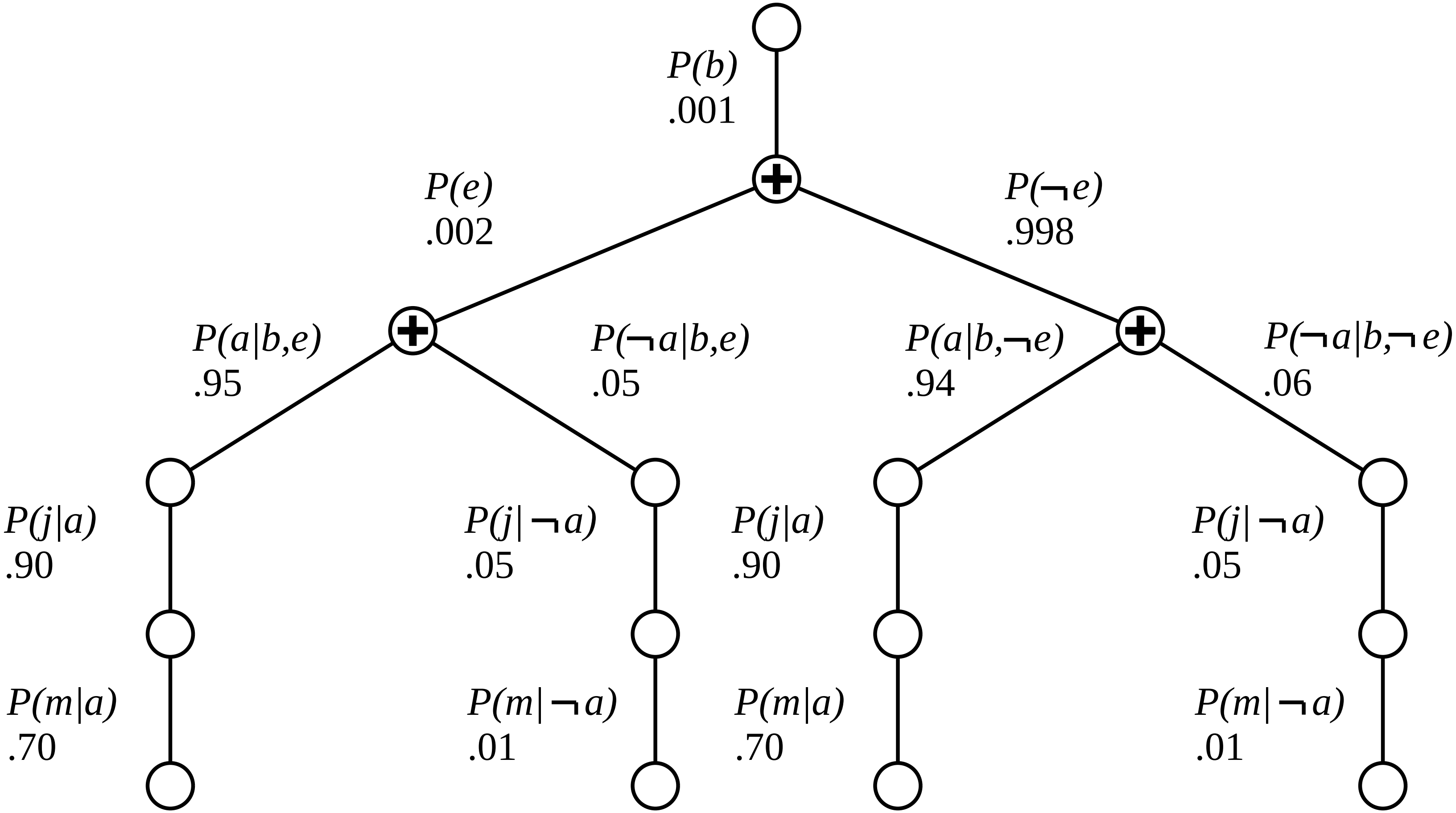
```
function Enumeration-Ask(X, e, bn) returns a distribution over X
   inputs: X,  the query variable
           e, observed values for variables E
           bn, a Bayesian network with variables {X} ∪ E ∪ Y

   Q(X) ← a distribution over X, initially empty
   for each value x_i of X do
      extend e with value x_i for X
      Q(x_i) ← Enumerate-All(Vars[bn], e)
   return Normalize(Q(X))
```

```
function Enumerate-All(vars, e) returns a real number
   if Empty?(vars) then return 1.0
   Y ← First(vars)
   if Y has value y in e
      then return P(y | Pa(Y)) × Enumerate-All(Rest(vars), e)
      else return Σ_y P(y | Pa(Y)) × Enumerate-All(Rest(vars), e_y)
           where e_y is e extended with Y = y
```

Albero di valutazione

L'enumerazione è inefficiente: calcoli ripetuti
p.e., calcola $P(j|a)P(m|a)$ per ogni valore di e



Inferenza tramite eliminazione di variabile

Eliminazione di variabile: effettuare le somme da destra a sinista, memorizzare i risultati intermedi (fattori) per evitare di ricalcolarli

$$
\begin{aligned}
&\mathbf{P}(B|j,m) \\
&\quad = \alpha\, \underbrace{\mathbf{P}(B)}_{B} \textstyle\sum_e \underbrace{P(e)}_{E} \textstyle\sum_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M} \\
&\quad = \alpha \mathbf{P}(B) \textstyle\sum_e P(e) \textstyle\sum_a \mathbf{P}(a|B,e) P(j|a) f_M(a) \\
&\quad = \alpha \mathbf{P}(B) \textstyle\sum_e P(e) \textstyle\sum_a \mathbf{P}(a|B,e) f_J(a) f_M(a) \\
&\quad = \alpha \mathbf{P}(B) \textstyle\sum_e P(e) \textstyle\sum_a f_A(a,b,e) f_J(a) f_M(a) \\
&\quad = \alpha \mathbf{P}(B) \textstyle\sum_e P(e) f_{\bar{A}JM}(b,e) \text{ (elimina } A\text{)} \\
&\quad = \alpha \mathbf{P}(B) f_{\bar{E}\bar{A}JM}(b) \text{ (elimina } E\text{)} \\
&\quad = \alpha f_B(b) \times f_{\bar{E}\bar{A}JM}(b)
\end{aligned}
$$

Eliminazione di variabile: operazioni base

Eliminare una variabile da un prodotto di fattori:

1. muovere i fattori costanti al di fuori della somma
2. aggiungere le sottomatrici al prodotto "pointwise" dei fattori rimanenti

$$\Sigma_x f_1 \times \cdots \times f_k = f_1 \times \cdots \times f_i \, \Sigma_x \, f_{i+1} \times \cdots \times f_k = f_1 \times \cdots \times f_i \times f_{\bar{X}}$$

assumendo che $f_1, \ldots, f_i$ non dipendano da X

Prodotto pointwise di fattori f_1 e f_2:

$$f_1(x_1, \ldots, x_j, y_1, \ldots, y_k) \times f_2(y_1, \ldots, y_k, z_1, \ldots, z_l)$$
$$= f(x_1, \ldots, x_j, y_1, \ldots, y_k, z_1, \ldots, z_l)$$

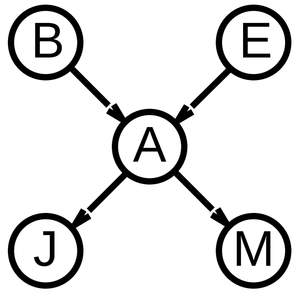
P.e., $f_1(a, b) \times f_2(b, c) = f(a, b, c)$

Algoritmo di eliminazione di variabile

```
function ELIMINATION-ASK(X, e, bn) returns a distribution over X
   inputs: X, the query variable
           e, evidence specified as an event
           bn, a belief network specifying joint distribution P(X_1, ..., X_n)

   factors ← [ ]; vars ← REVERSE(VARS[bn])
   for each var in vars do
       factors ← [MAKE-FACTOR(var, e)|factors]
       if var is a hidden variable then factors ← SUM-OUT(var, factors)
   return NORMALIZE(POINTWISE-PRODUCT(factors))
```

Variabili irrilevanti



Consideriamo la query$P(JohnCalls|Burglary = true)$

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

La somma su m è uguale a 1; M è **irrilevante** per la query

Thm 1: Y è irrilevante a meno che $Y \in Ancestors(\{X\} \cup \mathbf{E})$

Qui, $X = JohnCalls$, $\mathbf{E} = \{Burglary\}$, e
$Ancestors(\{X\} \cup \mathbf{E}) = \{Alarm, Earthquake\}$
quindi M è irrilevante

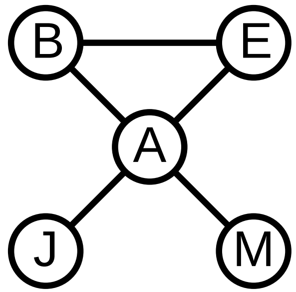
(Confrontare con backward chaining a partire dalla query in KB con clausole di Horn)

Variabili irrilevanti

Defn: grafo moralizzato di una rete bayesiana: sposare tutti i genitori ed eliminare la direzione degli archi

Defn: $\mathbf{F}$ è m-separato da $\mathbf{G}$ tramite $\mathbf{H}$ sse è separato tramite $\mathbf{H}$ nel grafo moralizzato

Thm 2: Y è irrilevante se m-separato da X tramite $\mathbf{E}$



Per $P(JohnCalls|Alarm = true)$, sia $Burglary$ che $Earthquake$ sono irrilevanti

Complessità dell'inferenza esatta

Reti singolarmente connesse (o polytree):
- ogni coppia di nodi è connessa da al più un cammino (non diretto)
- il costo in tempo e spazio della eliminazione di variabile è $O(d^k n)$

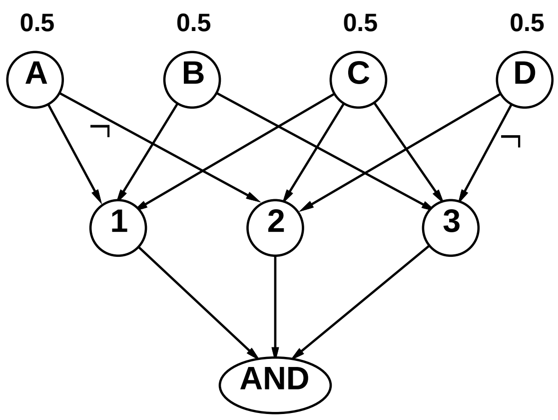
Reti connesse più che singolarmente:
- possibile ridurre 3SAT alla inferenza esatta $\Rightarrow$ NP-hard
- equivalente a modelli 3SAT con **conteggio** (del numero di soluzioni)

$\Rightarrow$ #P-complete

1. A v B v C

2. C v D v ¬A

3. B v C v ¬D



Inferenza Approssimata: metodi Monte Carlo "simulano" stocasticamente la rete