Reti Bayesiane I

IALab A.A. 2018/2019

Sintassi e Semantica

Bayesian networks

A simple, graphical notation for conditional independence assertions
and hence for compact specification of full joint distributions

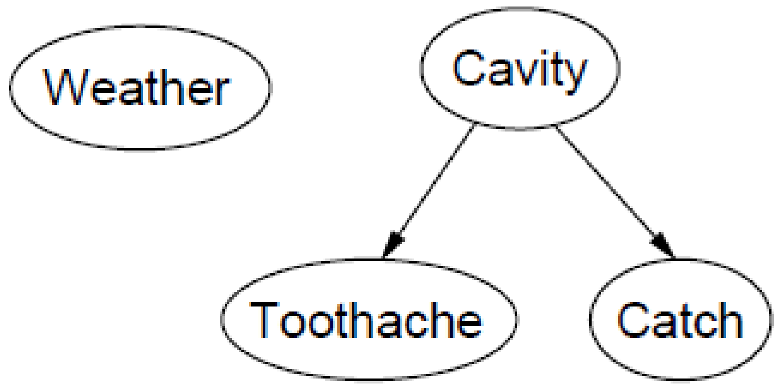
Syntax:

a set of nodes, one per variable
a directed, acyclic graph (link $\approx$ "directly influences")
a conditional distribution for each node given its parents:
$\mathbf{P}(X_i|Parents(X_i))$

In the simplest case, conditional distribution represented as
a conditional probability table (CPT) giving the
distribution over X_i for each combination of parent values

Example

Topology of network encodes conditional independence assertions:



$Weather$ is independent of the other variables

$Toothache$ and $Catch$ are conditionally independent given $Cavity$

Example

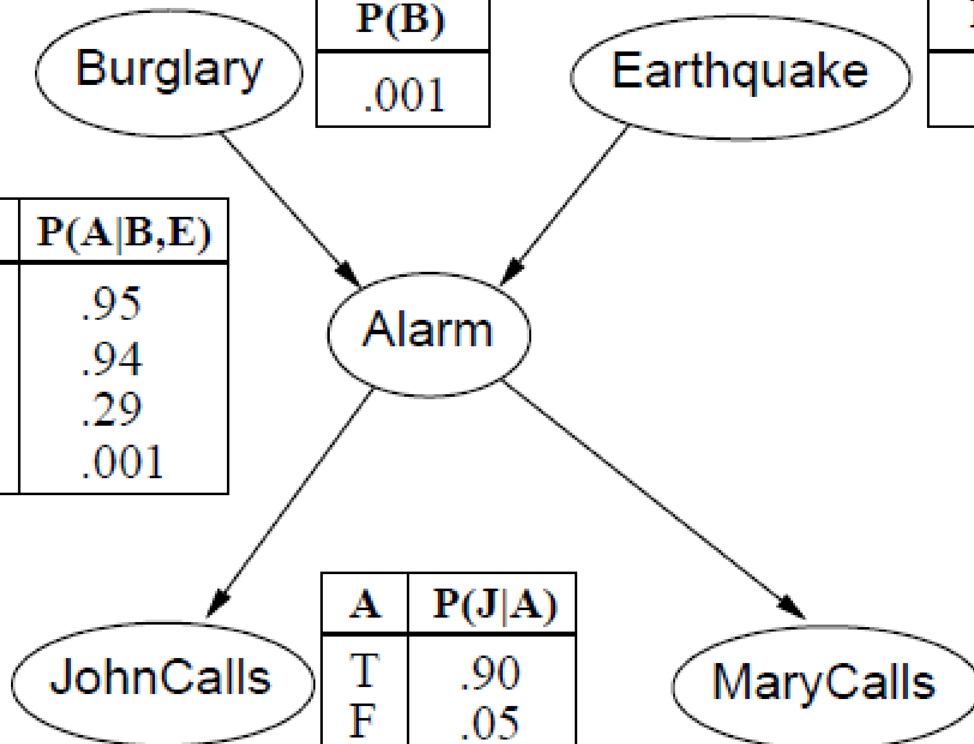
I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: *Burglar*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
Network topology reflects "causal" knowledge:

- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

non obbligatorio

Example contd.



P(B)
.001

P(E)
.002

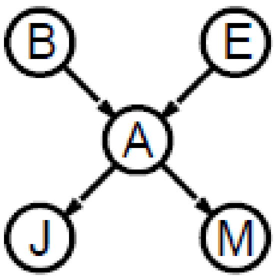
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

Compactness

A CPT for Boolean X_i with k Boolean parents has
2^k rows for the combinations of parent values



Each row requires one number p for $X_i = true$
(the number for $X_i = false$ is just $1 - p$)

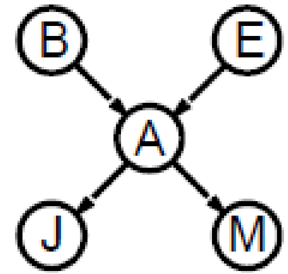
If each variable has no more than k parents,
the complete network requires $O(n \cdot 2^k)$ numbers

I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution

For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)

Global semantics

"Global" semantics defines the full joint distribution as the product of the local conditional distributions:



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

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

$$= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$
$$\approx 0.00063$$

Global Semantics VS Chain Rule

La Distribuzione Congiunta per l'esempio dell'Allarme è un'unica equazione:

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

(*32* equazioni su combinazioni complete di valori)

Con la *Chain Rule*:

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

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

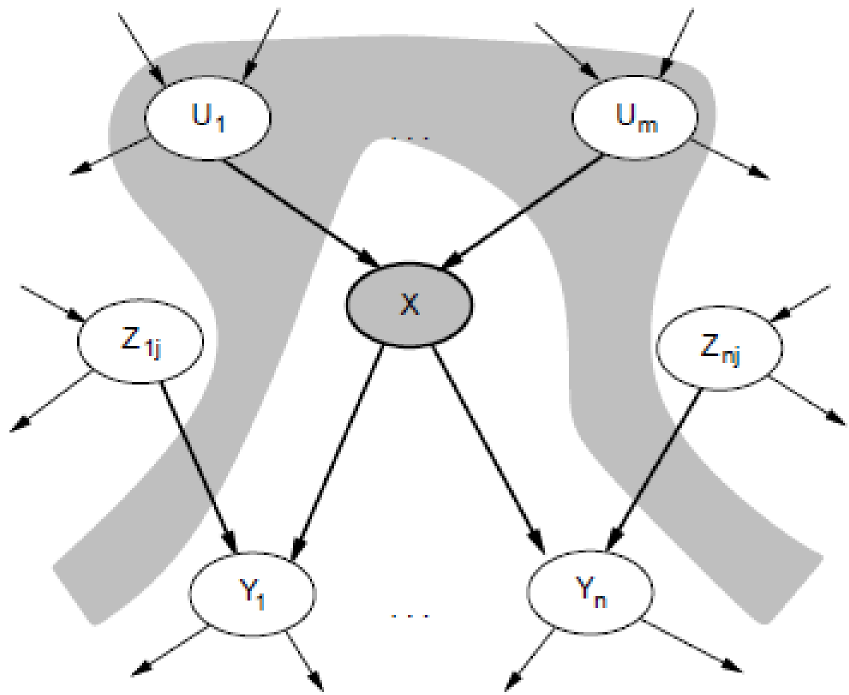
$$\mathbf{P}(J,M,A)\mathbf{P}(B \mid J,M,A)\mathbf{P}(E \mid J,M,A,B) =$$

$$\ldots$$

$$\mathbf{P}(J)P(M \mid J)\mathbf{P}(A \mid J,M)\mathbf{P}(B \mid J,M,A)\underbrace{\mathbf{P}(E \mid J,M,A,B)}_{\text{fattore più grande}}$$

Local semantics

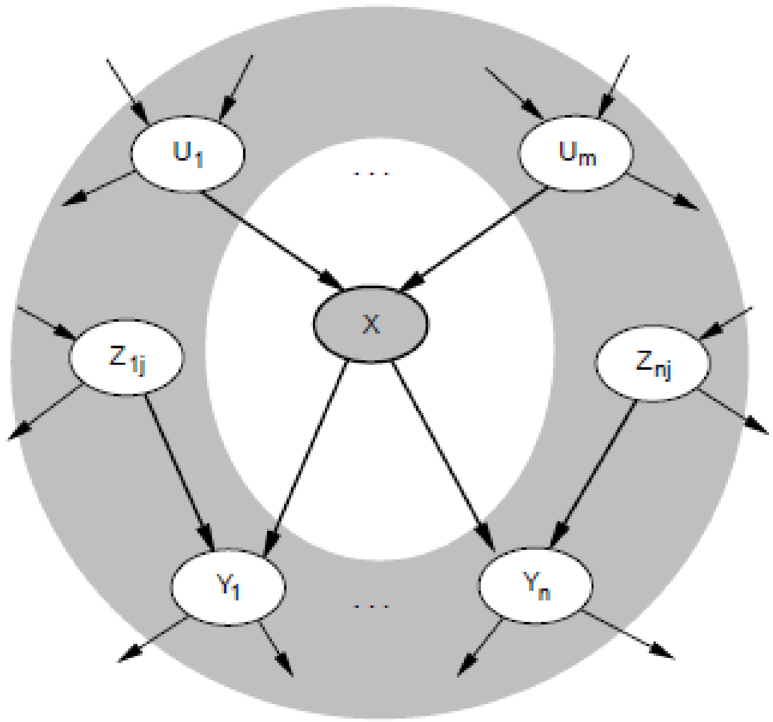
Local semantics: each node is conditionally independent of its nondescendants given its parents



Theorem: Local semantics $\Leftrightarrow$ global semantics

Markov blanket

Each node is conditionally independent of all others given its
Markov blanket: parents + children + children's parents



equivalente a Global/Local semantics

Constructing Bayesian networks

Need a method such that a series of locally testable assertions of conditional independence guarantees the required global semantics

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to n

   add X_i to the network

   select parents from $X_1, \ldots, X_{i-1}$ such that

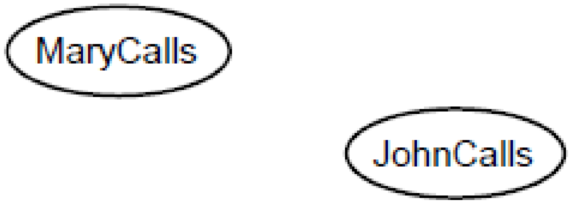
   $\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_1, \ldots, X_{i-1})$

This choice of parents guarantees the global semantics:

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

Example

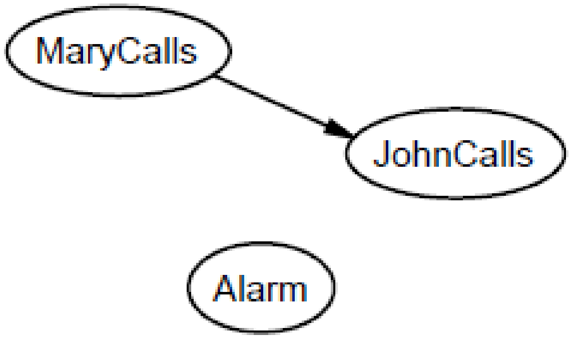
Suppose we choose the ordering M, J, A, B, E



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

Example

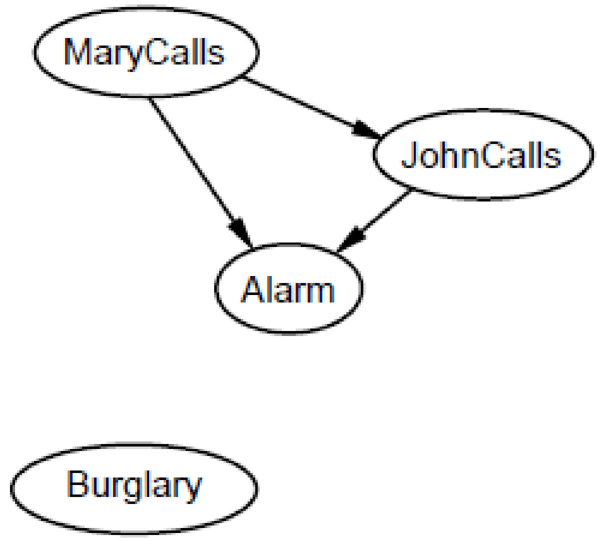
Suppose we choose the ordering M, J, A, B, E



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

Example

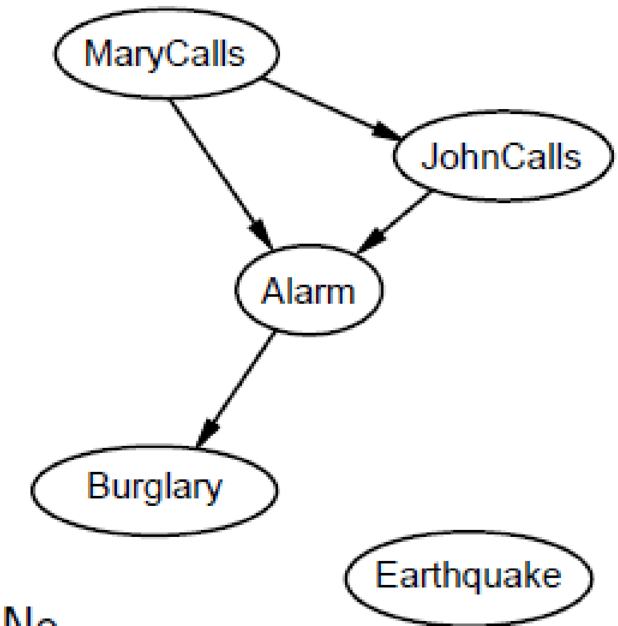
Suppose we choose the ordering M, J, A, B, E



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

Example

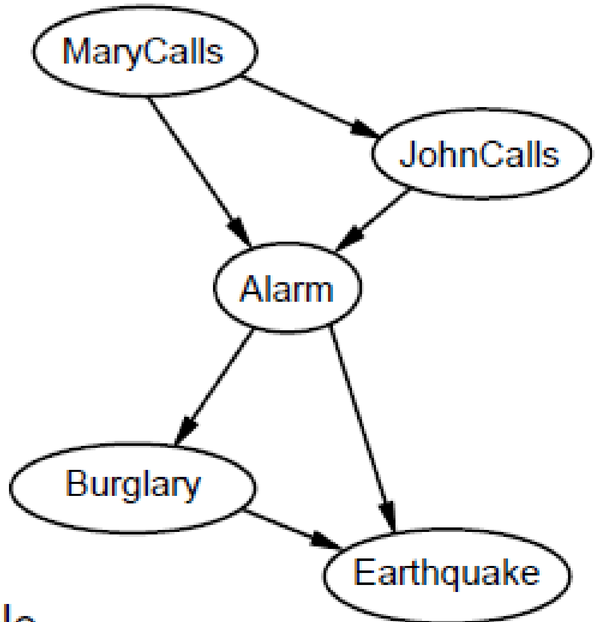
Suppose we choose the ordering M, J, A, B, E



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

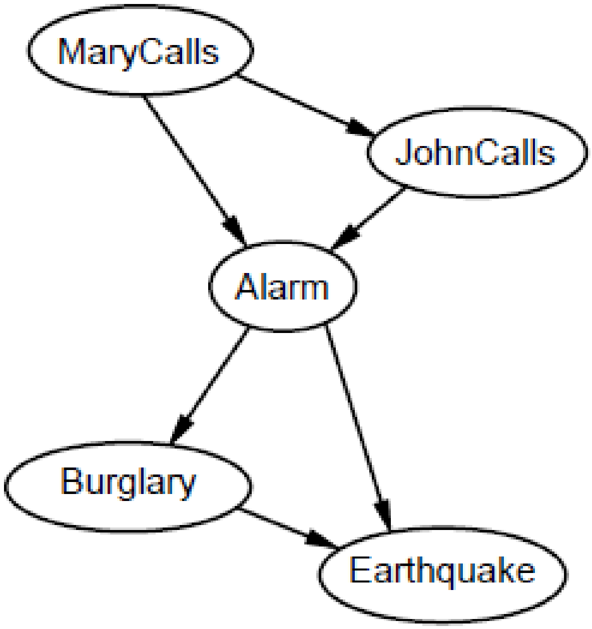
Example

Suppose we choose the ordering 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

Example contd.



Deciding conditional independence is hard in noncausal directions

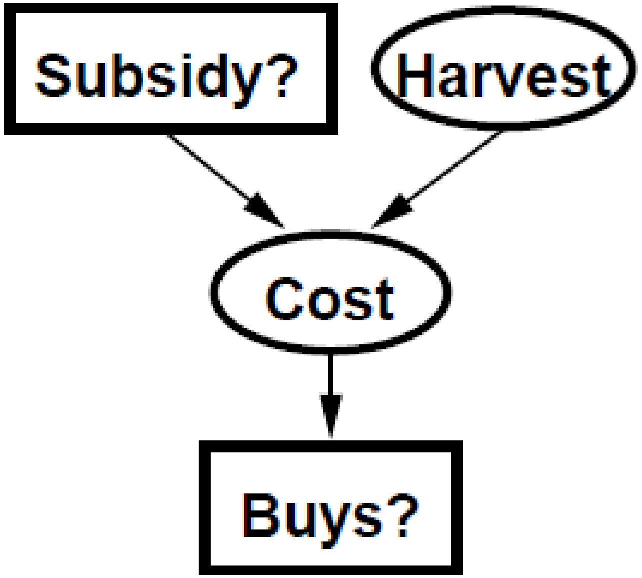
(Causal models and conditional independence seem hardwired for humans!)

Assessing conditional probabilities is hard in noncausal directions

Network is less compact: $1+2+4+2+4=13$ numbers needed

Hybrid (discrete+continuous) networks

Discrete (*Subsidy?* and *Buys?*); continuous (*Harvest* and *Cost*)



Option 1: discretization—possibly large errors, large CPTs
Option 2: finitely parameterized canonical families

1) Continuous variable, discrete+continuous parents (e.g., *Cost*)
2) Discrete variable, continuous parents (e.g., *Buys?*)

Continuous child variables

Need one conditional density function for child variable given continuous parents, for each possible assignment to discrete parents

funzione di densità di probabilità (fdp) condizionale

Most common is the linear Gaussian model, e.g.,:

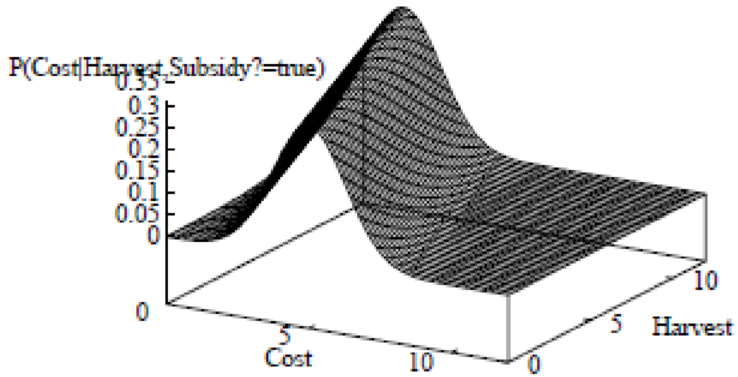
$$
\begin{aligned}
&P(Cost = c \mid Harvest = h, Subsidy? = true) \\
&= N(a_t h + b_t, \sigma_t)(c) \\
&= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c - (a_t h + b_t)}{\sigma_t}\right)^2\right)
\end{aligned}
$$

Mean $Cost$ varies linearly with $Harvest$, variance is fixed

Linear variation is unreasonable over the full range
but works OK if the likely range of $Harvest$ is narrow

"piecewise linear"

Continuous child variables



All-continuous network with LG distributions

⇒ full joint distribution is a multivariate Gaussian

Discrete+continuous LG network is a conditional Gaussian network i.e., a multivariate Gaussian over all continuous variables for each combination of discrete variable values

Discrete variable w/ continuous parents

Probability of $Buys?$ given $Cost$ should be a "soft" threshold:

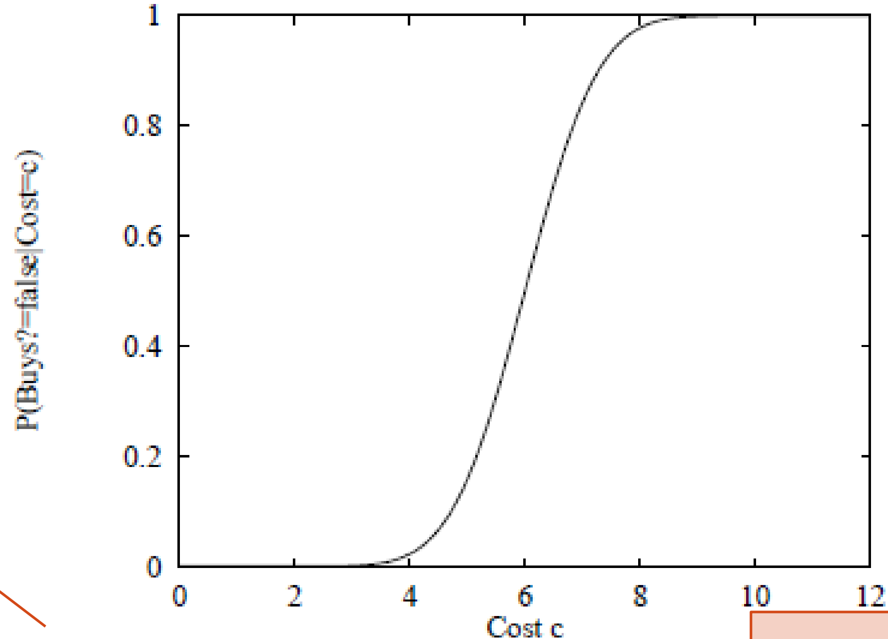


grafico: B?=false
formula: B?=true

Probit distribution uses integral of Gaussian:

$$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx$$

$$P(Buys? = true \mid Cost = c) = \Phi((-c + \mu)/\sigma)$$

Funzione di Distribuzione Cumulativa (FDC) di N(0,1)

Inference tasks

Simple queries: compute posterior marginal $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$
e.g., $P(NoGas|Gauge=empty, Lights=on, Starts=false)$

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

Optimal decisions: decision networks include utility information;
probabilistic inference required for $P(outcome|action, evidence)$

Value of information: which evidence to seek next?

Sensitivity analysis: which probability values are most critical?

Explanation: why do I need a new starter motor?

molti modi di intendere questa parola

Inference Tasks

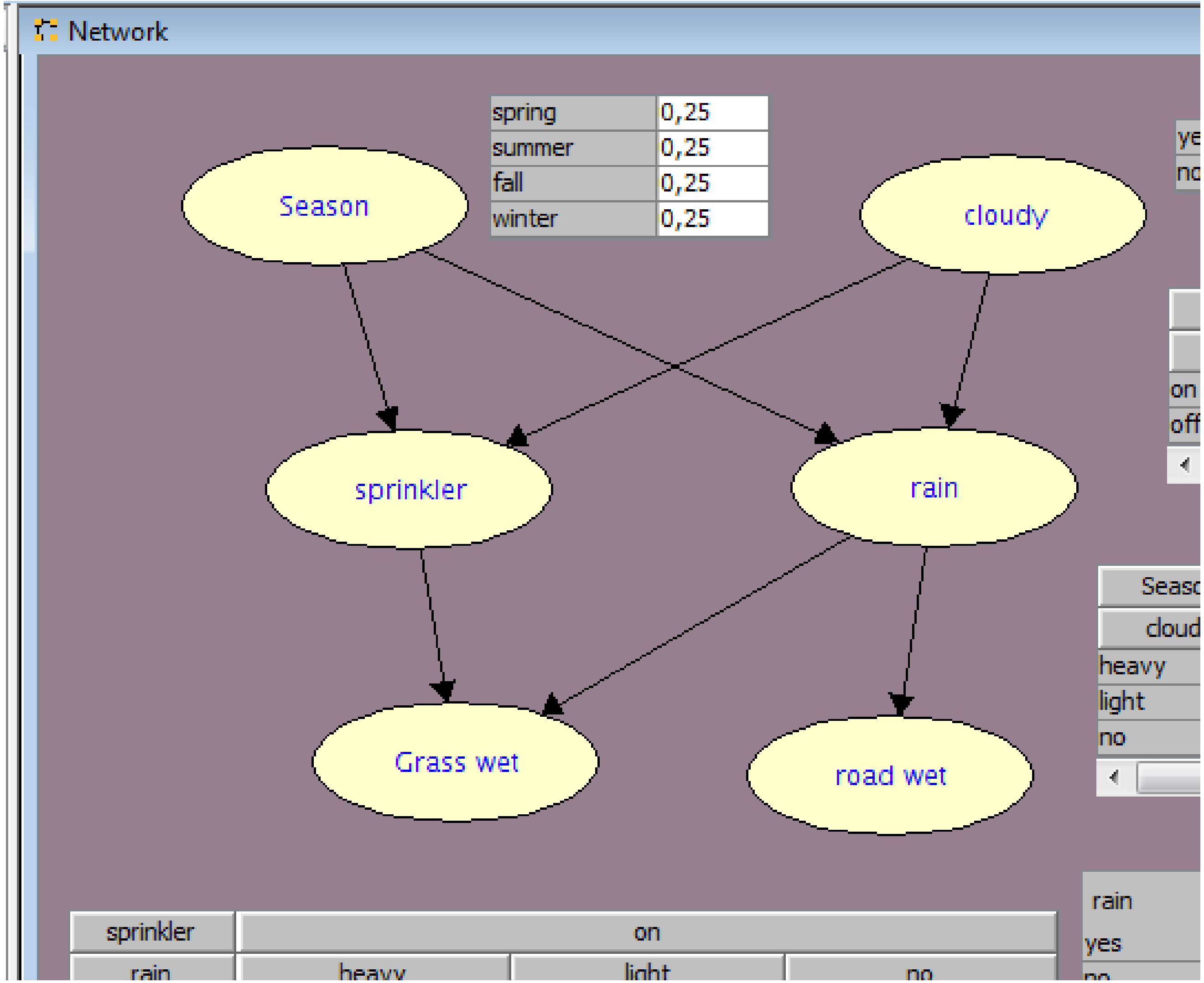
- **Most Probable Explanation** (MPE): qual è la istanza **x** più probabile di tutte le variabili **X** data l'evidenza **e**?

$$MPE(\boldsymbol{e}) = argmax_{\boldsymbol{x}} P(\boldsymbol{x}, \boldsymbol{e})$$

- **Maximum a Posteriori Probability** (MAP): qual è la istanza **m** più probabile di un sottoinsieme **M**$\subseteq$**X** di variabili data l'evidenza **e**?

$$MAP(\boldsymbol{e}) = argmax_{\boldsymbol{m}} P(\boldsymbol{m}, \boldsymbol{e})$$

Esecizio su Sprinkler (Samlam)



Esecizio su Sprinkler (SamIam)

- creare la rete SprinklerPlus con SamIam
- se sappiamo che *(Grass wet = yes)* qual è la probabilità a posteriori di *Season*? Spiegare il risultato intuitivamente
- qual è la stagione in cui è più probabile che piova?
- se so che *Grass wet* è **yes** e che *Road wet* è **yes**, qual è la combinazione più probabile di valori delle variabili rimanenti? Spiegare risultato intuitivamente

[usare **shenoy-shafer** per calcolare **MPE**]

Esecizio su Sprinkler (Samlam)

- se so che *Grass wet* è **yes** e che *Road wet* è **yes**, qual è la combinazione più probabile di valori delle variabili *Season*, *Cloudy*, *Sprinkler*? Spiegare il risultato intuitivamente

[impostare le variabili-risultato per **MAP** usando il bottone della toolbar “Variable Selection” e poi selezionando la voce “map” dalla lista della seconda colonna]

Esecizio su Sprinkler (Samlam)

Season	P(Season)
Spring	0.25
Summer	0.25
Fall	0.25
Winter	0.25

Cloudy	P(Cloudy)
yes	0.4

Season	Cloudy	P(Sprinkler|…)=on
Spring	Yes	0.01
Summer	Yes	0.2
Fall	Yes	0.001
Winter	Yes	0.0
Spring	No	0.1
Summer	No	0.7
Fall	No	0.01
Winter	No	0

Esecizio su Sprinkler (Samlam)

Season	Cloudy	P(Rain | …)=heavy	P(Rain | …)=light
Spring	Yes	0.2	0.4
Summer	Yes	0.6	0.15
Fall	Yes	0.3	0.5
Winter	Yes	0.01	0.79
Spring	No	0	0.1
Summer	No	0.01	0.04
Fall	No	0.001	0.2
Winter	No	0	0.01

Esecizio su Sprinkler (Samlam)

Rain	P(RoadW | …)=T
Heavy	1
Light	0.9
No	0.01

Sprinkler	Rain	P(Gr assW | …)=T
On	Heavy	1.00
On	Light	0.99
On	No	0.95
Off	Heavy	0.999
Off	Light	0.8
Off	No	0.01

Cows Artificial Insemination

- modellare con una BN la seguente situazione [da Darwiche]:

A few weeks after artificial insemination of a cow, we have three possible tests to confirm pregnancy. The first is a scanning test which has a false positive of 1% and a false negative of 10%. The second is a blood test, which detects progesterone with a false positive of 10% and a false negative of 30%. The third test is a urine test, which also detects progesterone with a false positive of 10% and a false negative of 20%. The probability of a detectable progesterone level is 90% given pregnancy, and 1% given no pregnancy. The probability that insemination will impregnate a cow is 87%.