Intro ML
ML. 5. Probabilistic graphical models

DataLab CSIC

Objectives and schedule

Introduce key concepts about PGMs. Conditional independence. Representations: directed and undirected. Hints on computations and inference. Influence diagrams. Generative models, Gibbs sampling.

Contents

- Bayesian networks
- Conditional Independence
- Markov random fields
- Inference
- Influence diagrams

Bishop 8, Goodfellow et al 16

Lab

- Several lasb around probabilistic graphical models
  - Handling PGMs
  - Structuring PGMs
- Case. Risk factors for cardiovascular diseases by Chem Camacho

PGMs. Motivation

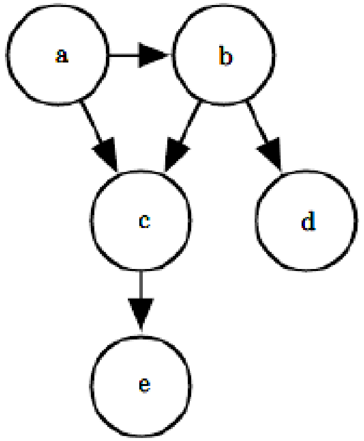
Motivation

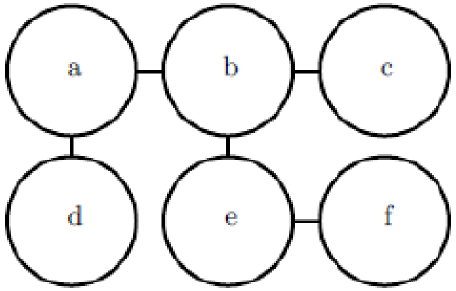
- Simple way to visualize structure of probabilistic models
- Designing and motivating new models
- Understanding properties like conditional independence
- Complex computations viewed through simple graphical manipulations
- Explainable and interpretable
- Classification, generation.
- Deep belief nets in deep learning

Concept

$$p(\mathbf{x}) = \prod_i p\left(\mathrm{x}_i \mid Pa_{\mathcal{G}}(\mathrm{x}_i)\right)$$

$$\tilde{p}(\mathbf{x}) = \Pi_{\mathcal{C}\in\mathcal{G}}\,\phi(\mathcal{C})$$





$$p(\mathrm{a},\mathrm{b},\mathrm{c},\mathrm{d},\mathrm{e}) = p(\mathrm{a})p(\mathrm{b}\mid\mathrm{a})p(\mathrm{c}\mid\mathrm{a},\mathrm{b})p(\mathrm{d}\mid\mathrm{b})p(\mathrm{e}\mid\mathrm{c})$$

$$p(\mathrm{a},\mathrm{b},\mathrm{c},\mathrm{d},\mathrm{e},\mathrm{f}) \quad \frac{1}{Z}\phi_{\mathrm{a},\mathrm{b}}(\mathrm{a},\mathrm{b})\phi_{\mathrm{b},\mathrm{c}}(\mathrm{b},\mathrm{c})\phi_{\mathrm{a},\mathrm{d}}(\mathrm{a},\mathrm{d})\phi_{\mathrm{b},\mathrm{e}}(\mathrm{b},\mathrm{e})\phi_{\mathrm{e},\mathrm{f}}(\mathrm{e},\mathrm{f})$$

Bayesian networks. Directed, Acyclic

Markov fields. Undirected

Probabilistic graphical models. Directed Bayesian networks

Directed PGMs

As basic tools for qualitative modelling of uncertainty use probabilistic influence diagrams a.k.a. causal networks, Bayesian networks, Belief networks,…. See the excellent

http://en.wikipedia.org/wiki/Bayesian_network

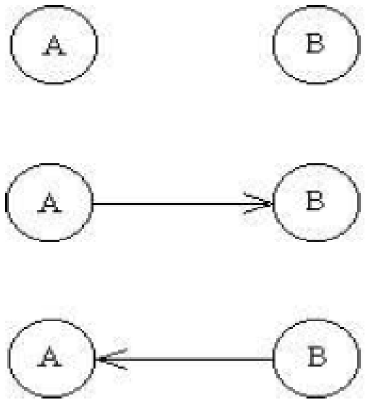
They are **influence diagrams** with chance nodes only. Qualitatively they describe a probabilistic model through

$$P(A1, A2, \ldots, An) = P(A1 \mid ant(A1)) \ldots P(An \mid ant(An))$$

where ant (Ai) are the antecessors of node Ai.

In what follows we see several PIDs and we need to indicate the entailed probabilistic model

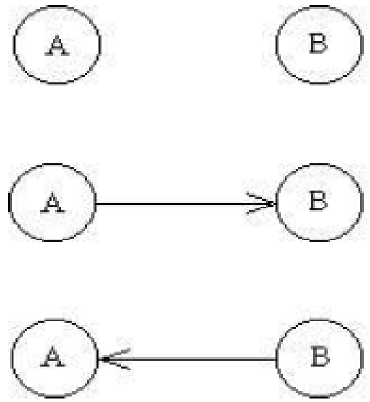
Probabilistic diagrams with two nodes



Before moving foreward, write the entailed probabilistic model

Probabilistic diagrams with two nodes

Model for P(A,B)



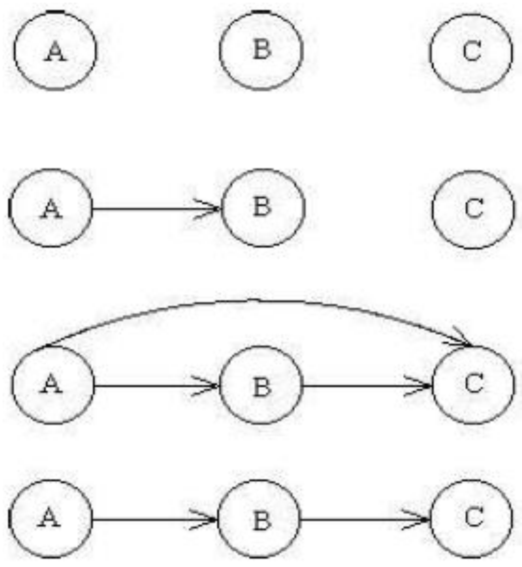
P(A)P(B)

P(A) P(B|A)

P(B) P(A|B)

First case, A and B are independent. We move from second to third, and viceversa, via Bayes formula

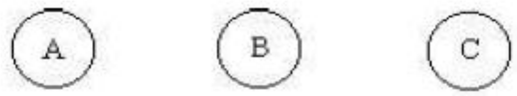
Probabilistic diagrams with three nodes



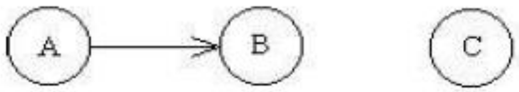
Before moving foreward, write the entailed probabilistic model

Probabilistic diagrams with three nodes

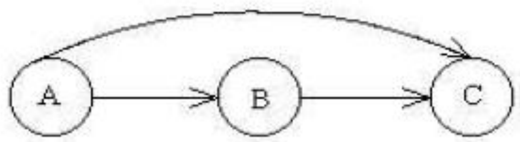
Model P(A ,B ,C)



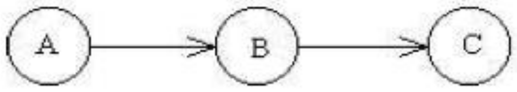
P(A)P(B)P(C)



P(A) P(B|A) P(C)



P(A)P(B|A)P(C|A,B)

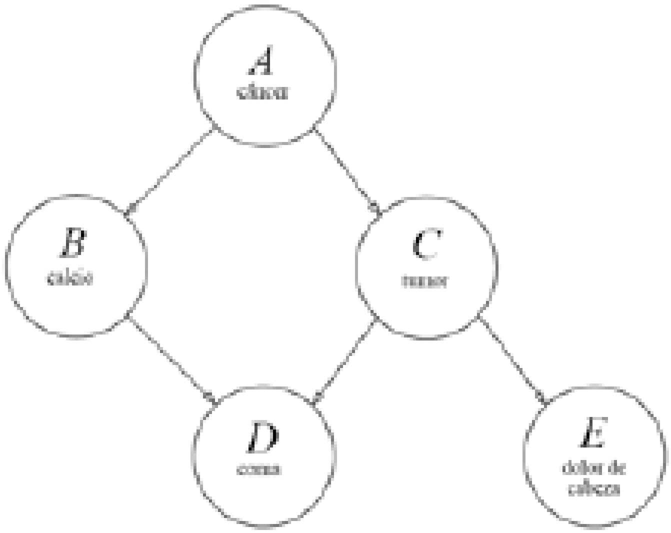


P(A)P(B|A)P(C|B)

First case, independence. Third case, A and C are conditionally independent given B.

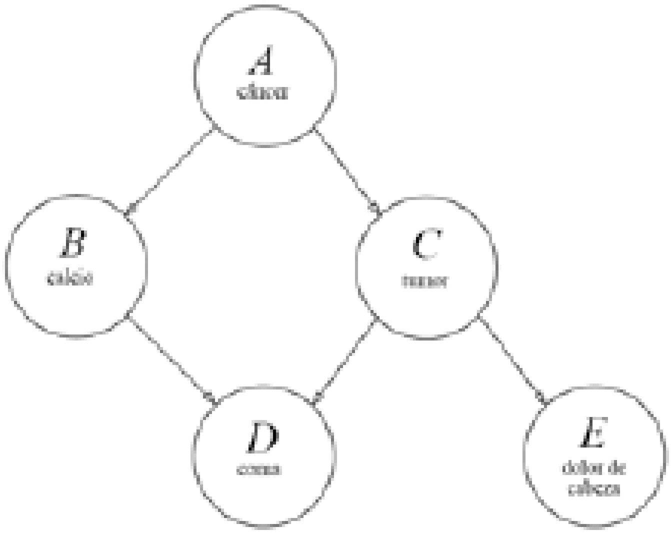
Read http://en.wikipedia.org/wiki/Conditional_independence

The hidden info



$$P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)P(E|C)$$

The hidden info



a	0.2

	a	$\overline{a}$
b	0.8	0.2

	a	$\overline{a}$
c	0.2	0.05

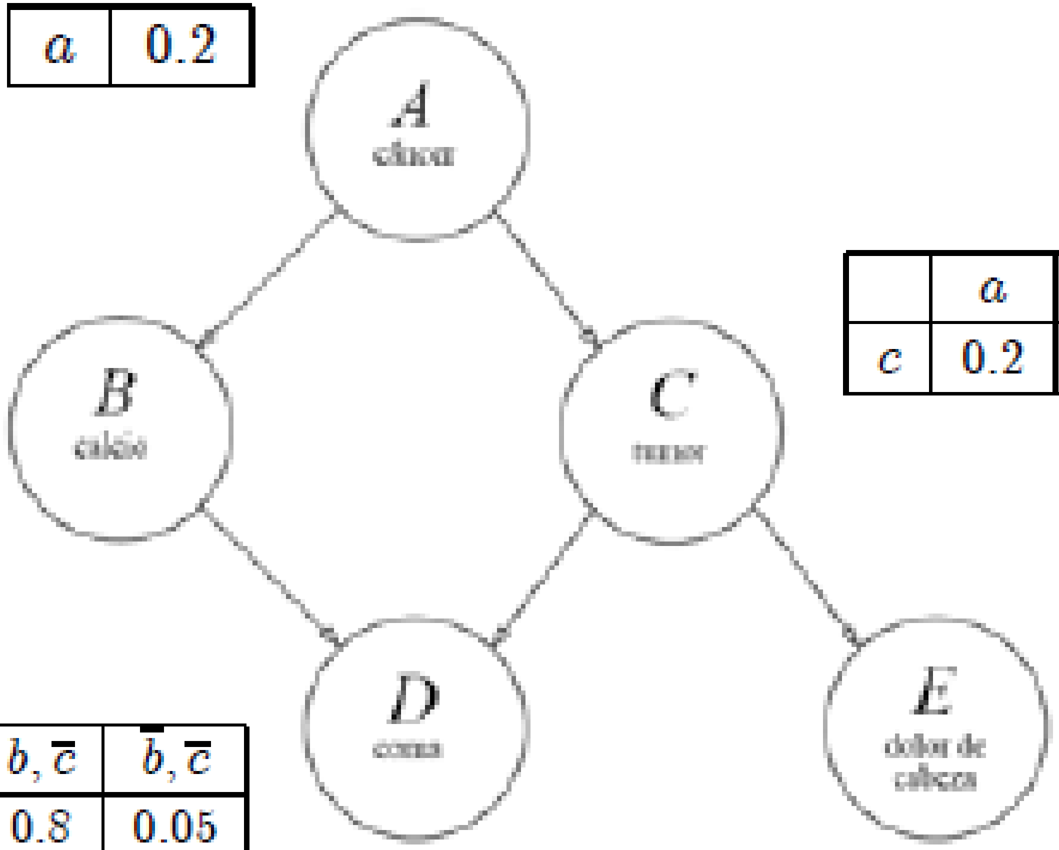
	b, c	$\overline{b}, c$	$b, \overline{c}$	$\overline{b}, \overline{c}$
d	0.8	0.8	0.8	0.05

	c	$\overline{c}$
e	0.8	0.6

$$P(A,B,C,D,E) = P(A)P(B|A)P(C|A)P(D|B,C)P(E|C)$$

The hidden info

a	0.2



	a	$\overline{a}$
b	0.8	0.2

	a	$\overline{a}$
c	0.2	0.05

	b, c	$\overline{b}, c$	$b, \overline{c}$	$\overline{b}, \overline{c}$
d	0.8	0.8	0.8	0.05

	c	$\overline{c}$
e	0.8	0.6

Conditional Independence I

A and B conditional independent given C if

$$p(A|B,C)=p(A|C)$$

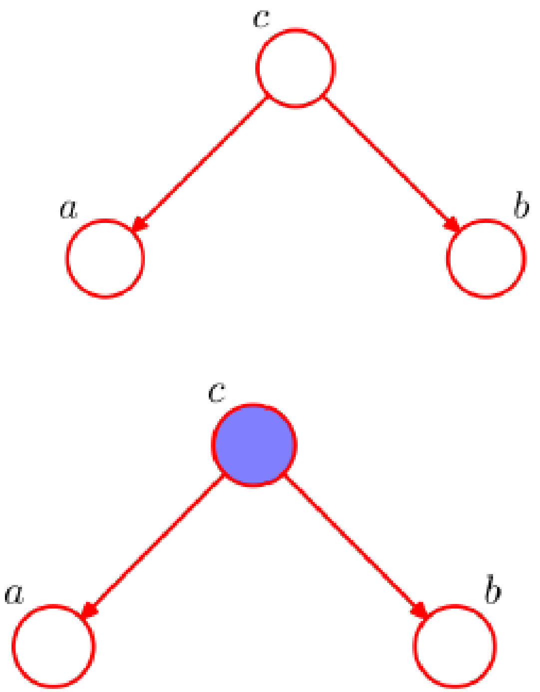
or

$$p(A,B|C) = p(A|C)\, p(B|C)$$

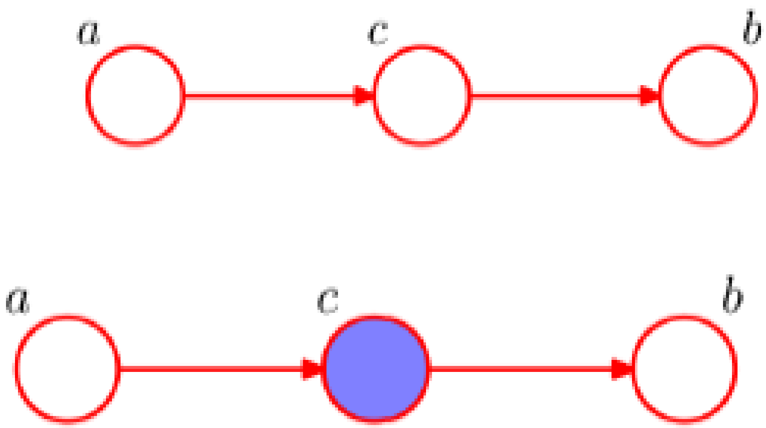
$$a \perp\!\!\!\perp b \mid c$$

Conditional Independence II

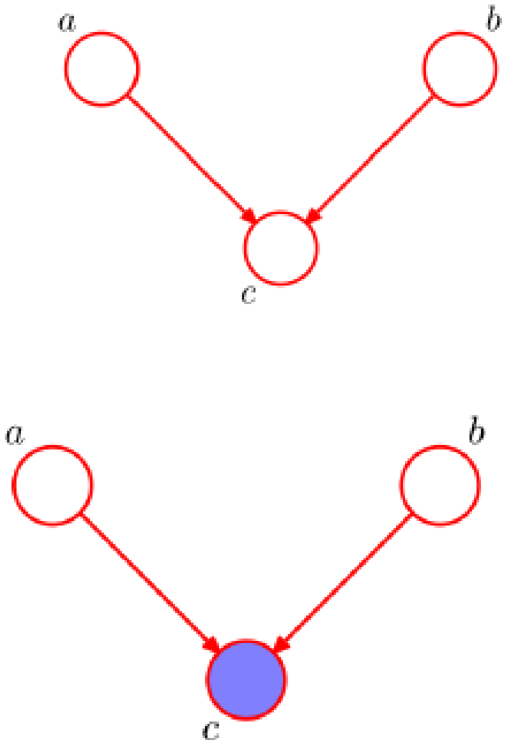
Tail to tail



Head to tail



Head to head



D-separation

In a DAG, let A, B and C be three different subsets of nodes

Are A and B c.i. given C?

Consider all trajectories between A and B. We say that a trajectory is blocked if it includes a node such that either

- All arrows on pathe find a head-to-tail or tail-to-tail at node and node in C, or
- All arrows find head-to-head at node, and is not in C (neither its descendants)

If all paths between A and B are blocked, then A and B are d-separated by C and are c.i. given C

Probabilistic diagrams. Asia

An example referring to lung diseases

A breathing condition (dyspnea) may be due to tuberculosis, lung cancer or bronchitis, none of them or several of them. A recent visit to Asia, increases the chances of tuberculosis, whereas smoking is a risk factor for lung cancer and bronchitis. The results of an X-ray may not discriminate between cancer and tuberculosis, as neither the presence or absence of dyspnea does.

Probabilistic diagrams

An example referring to lung diseases:

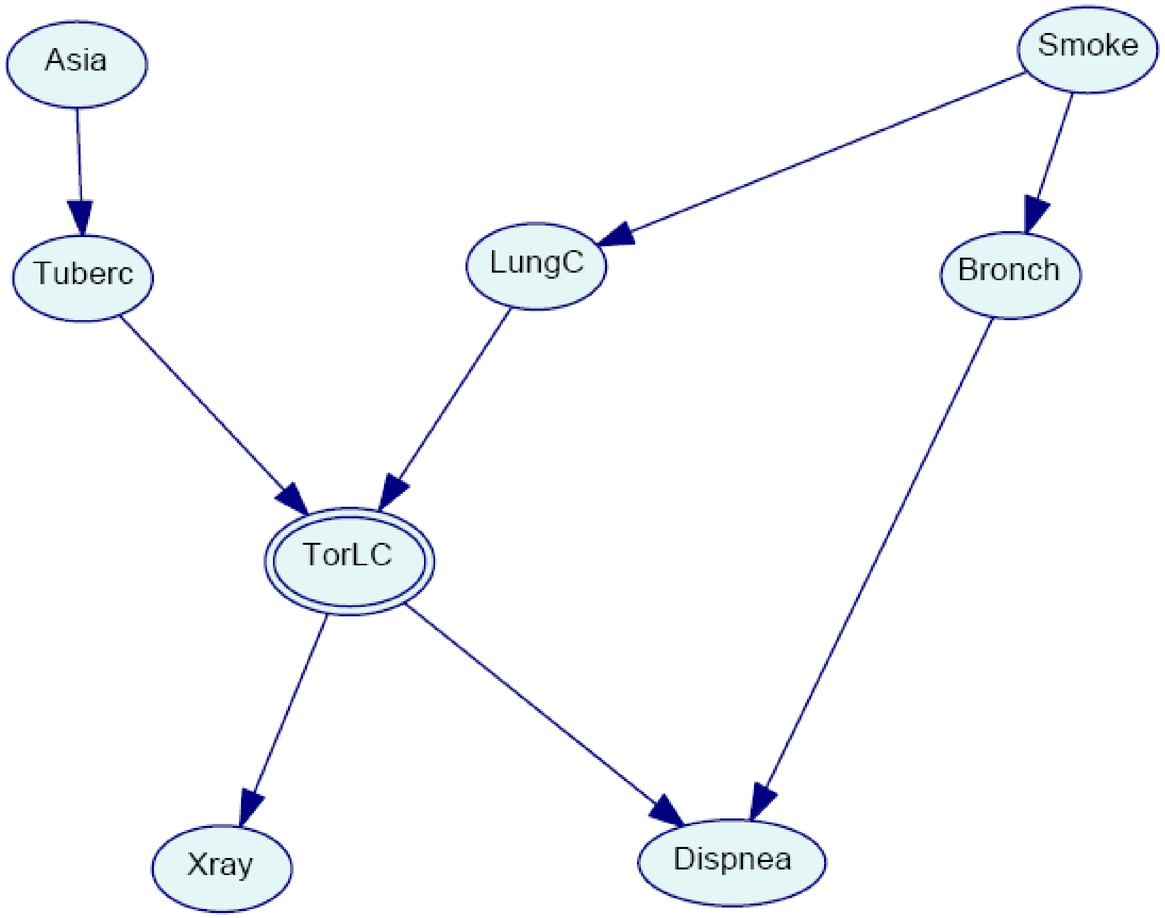
A breathing condition (dyspnea) may be due to tuberculosis, lung cancer or bronchitis, none of them or several of them. A recent visit to Asia, increases the chances of tuberculosis, whereas smoking is a risk factor for lung cancer and bronchitis. The results of an X-ray may not discriminate between cancer and tuberculosis, as neither the presence or absence of dyspnea does.

Probabilistic diagrams

An example referring to lung diseases

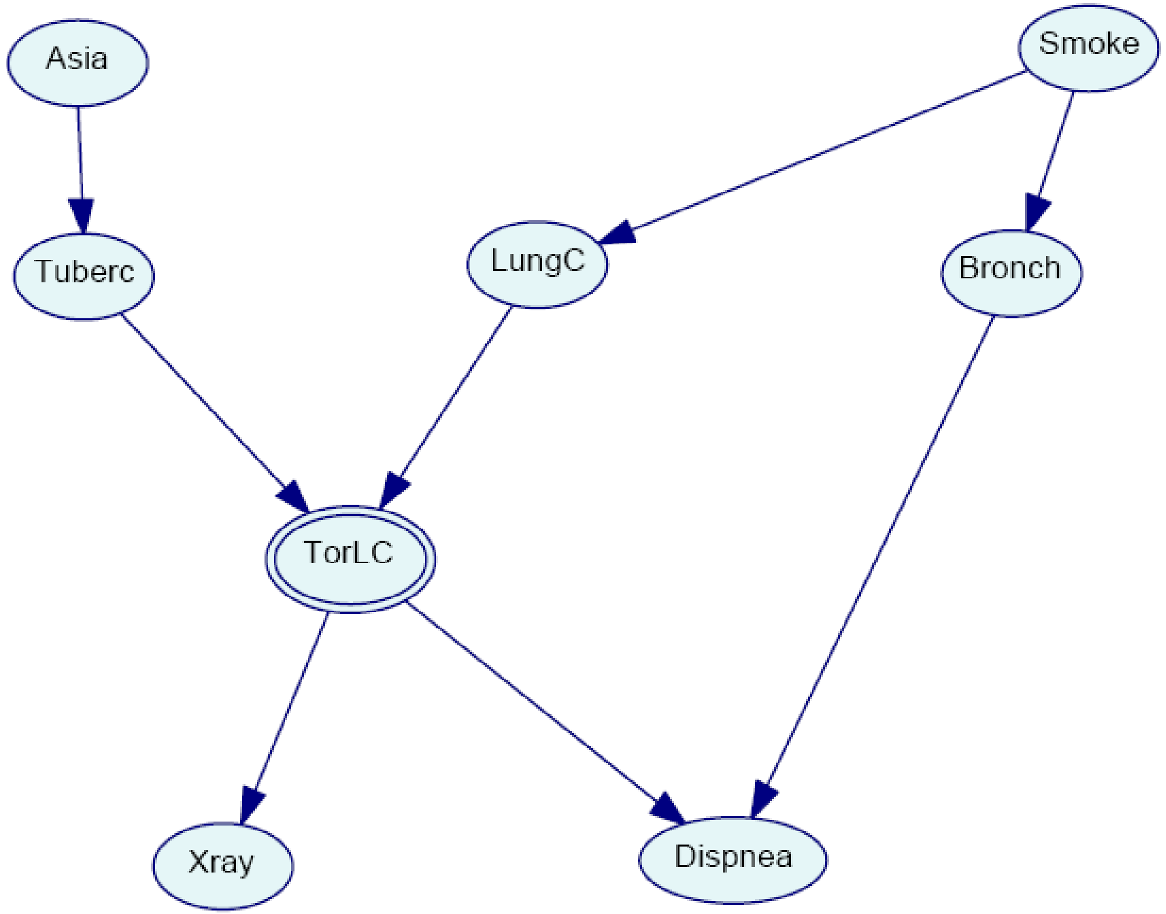
A breathing condition (dyspnea) may be due to tuberculosis, lung cancer or bronchitis, none of them or several of them. A recent visit to Asia, increases the chances of tuberculosis, whereas smoking is a risk factor for lung cancer and bronchitis. The results of an X-ray may not discriminate between cancer and tuberculosis, as neither the presence or absence of dyspnea does.

Probabilistic diagrams



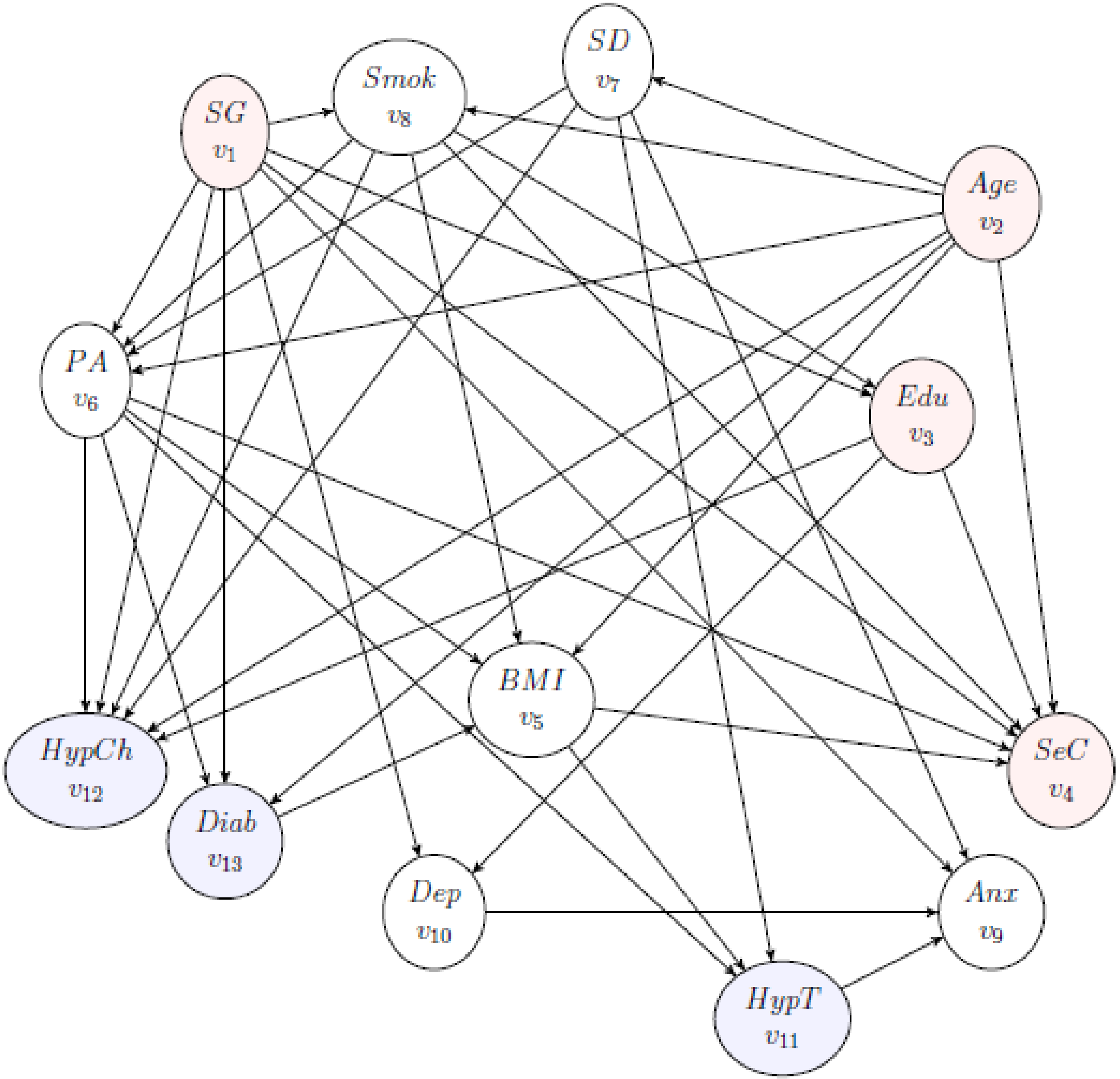
Provide the model

Probabilistic diagrams



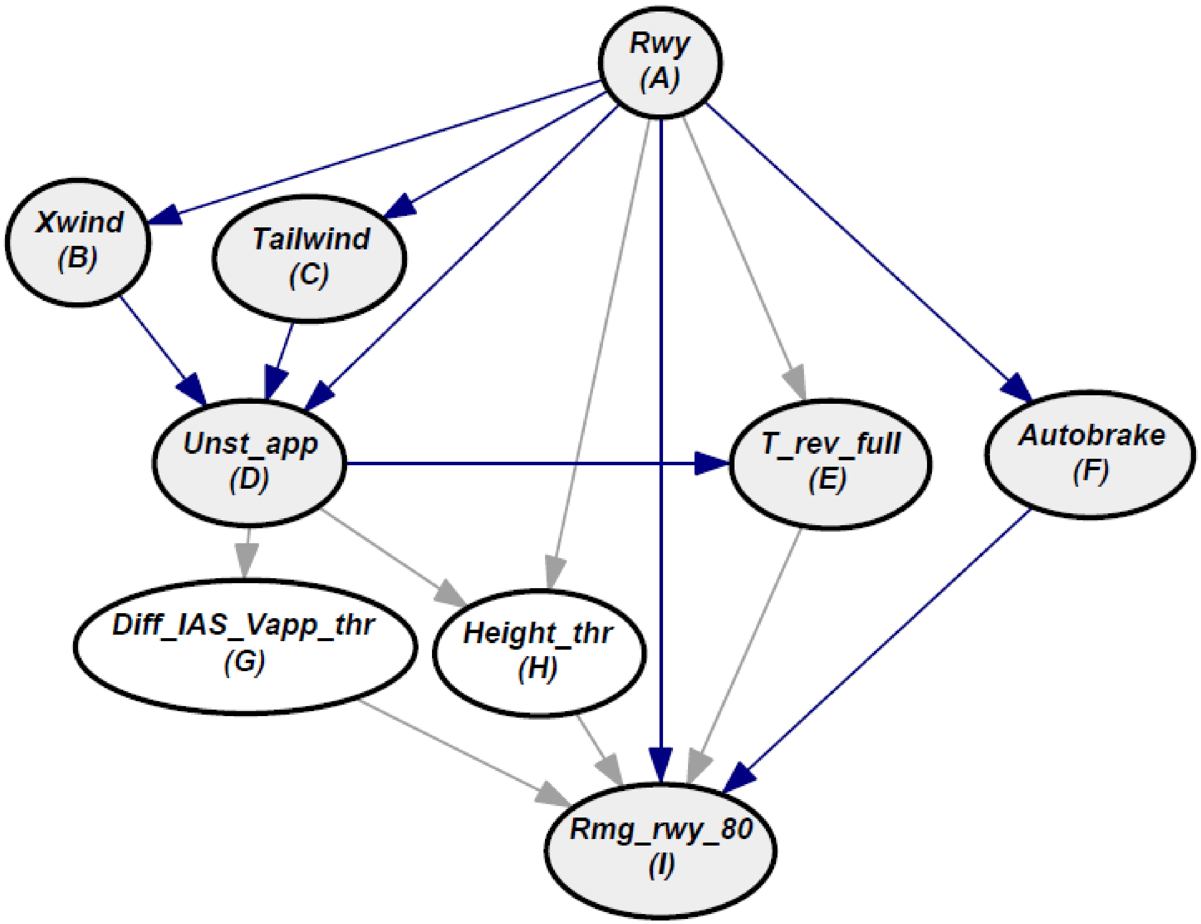
P(A,T,S,L,B,O,X,D) = P(A)P(T|A)P(S)P(L|S)P(B|S)P(0|T,L)P(X|O)P(D|O,B)

Hypertension



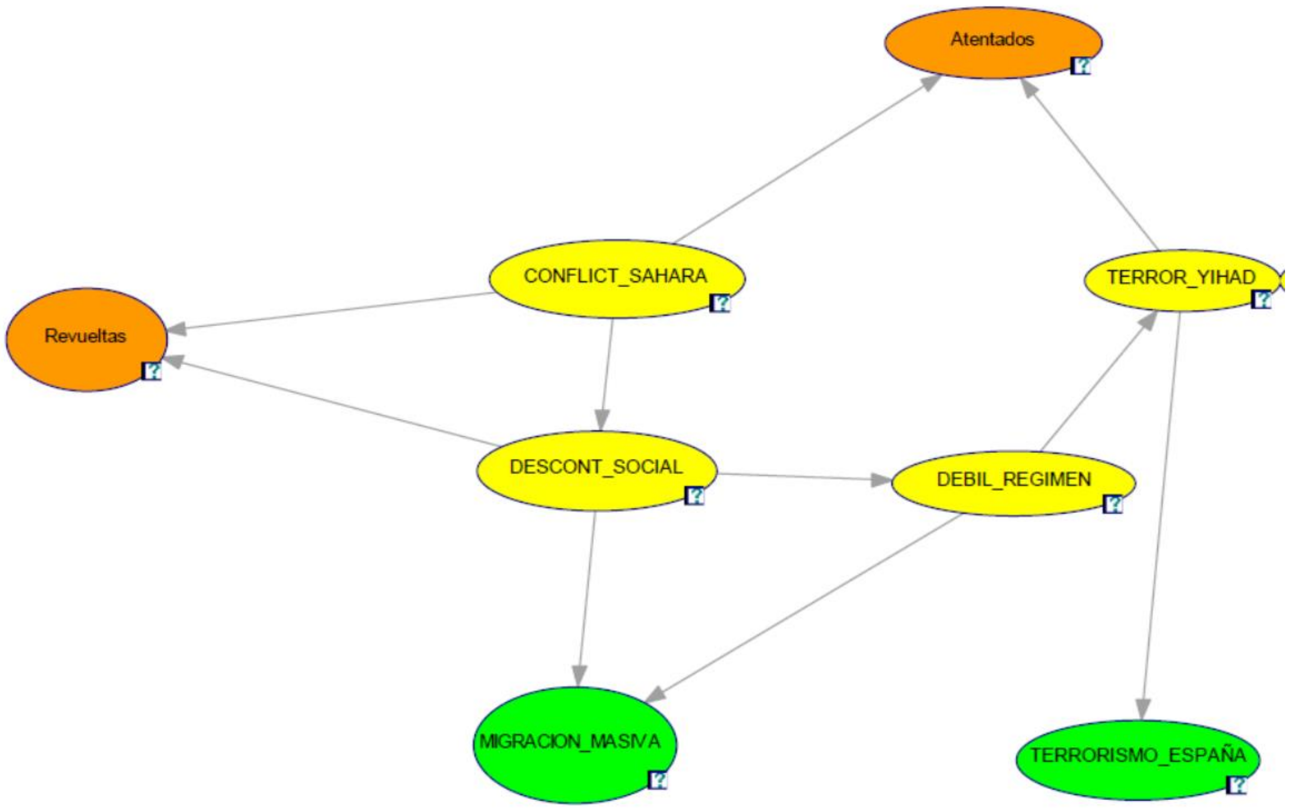
Build the probabilistic model

Runway excursions at airports



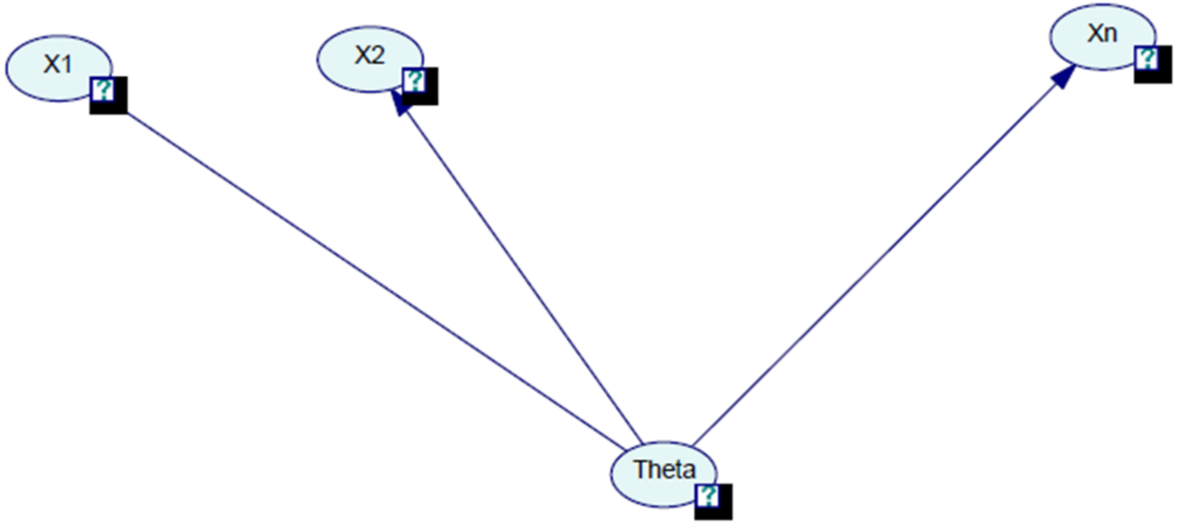
Build the probabilistic model

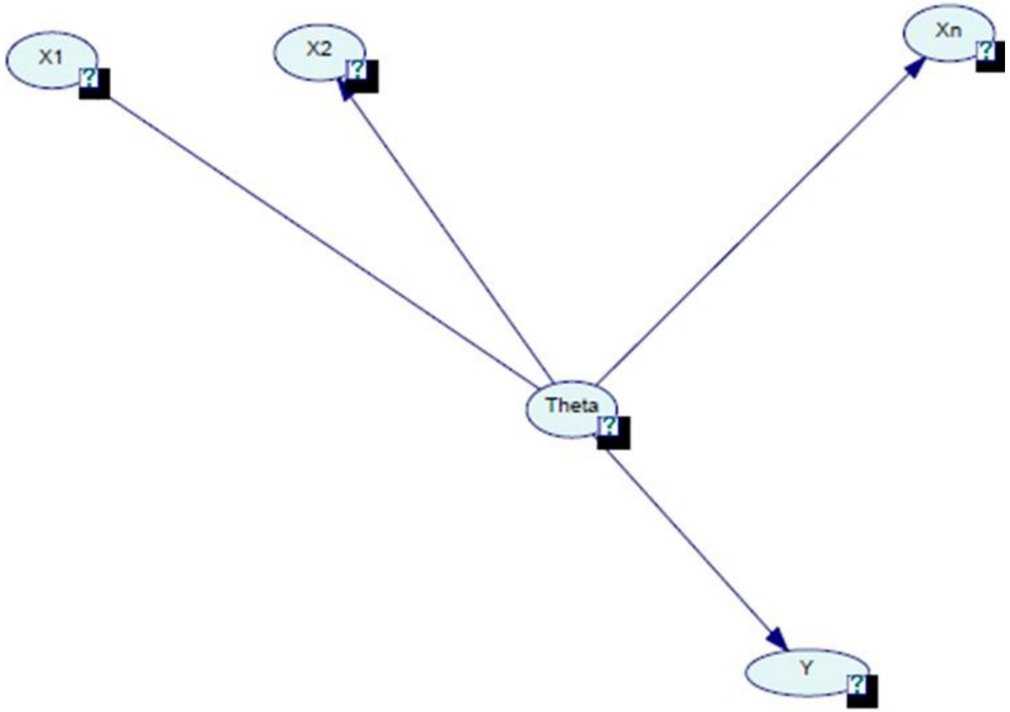
National security



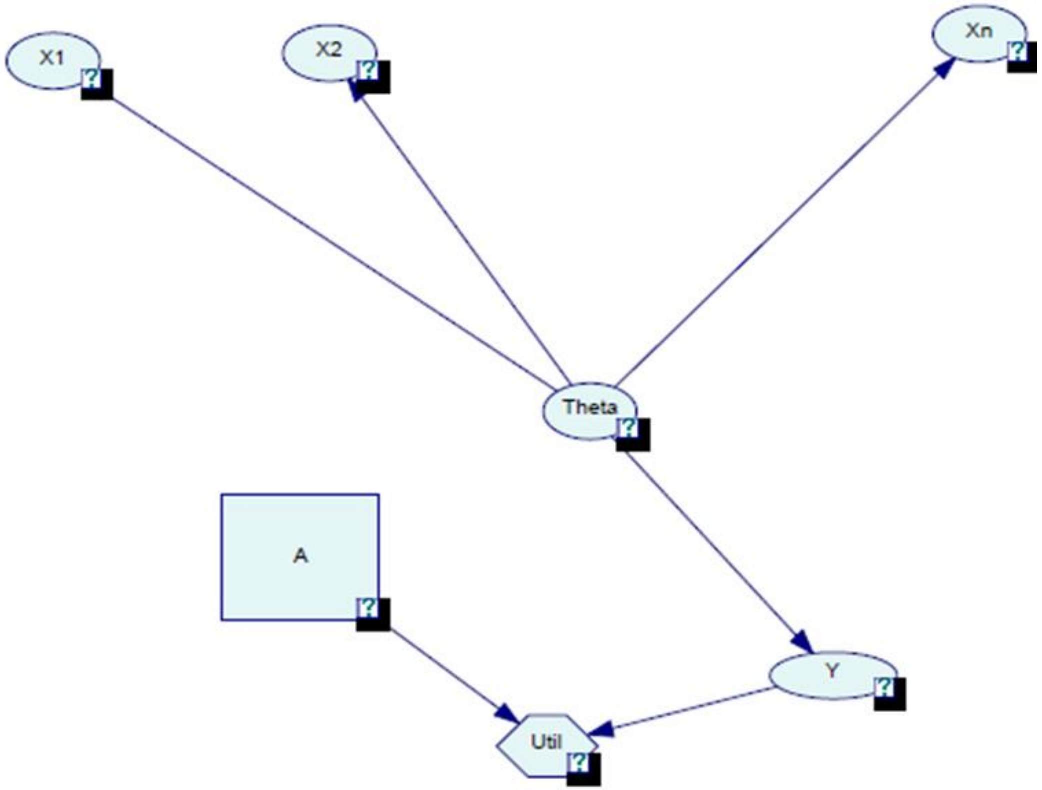
Build the probabilistic model

Statistical models as PGMs. Inference and Prediction

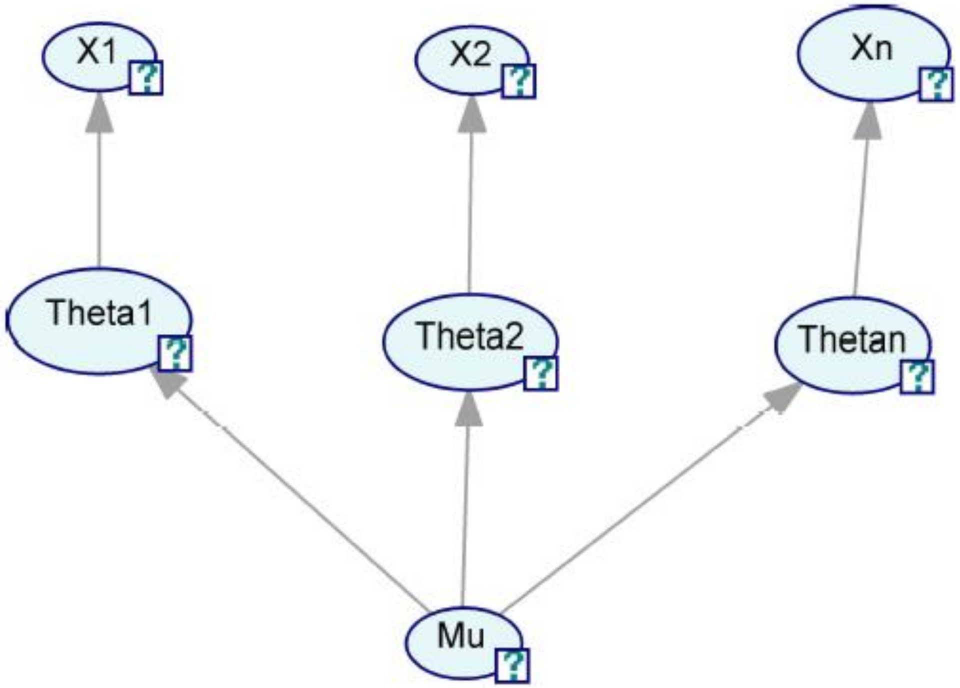




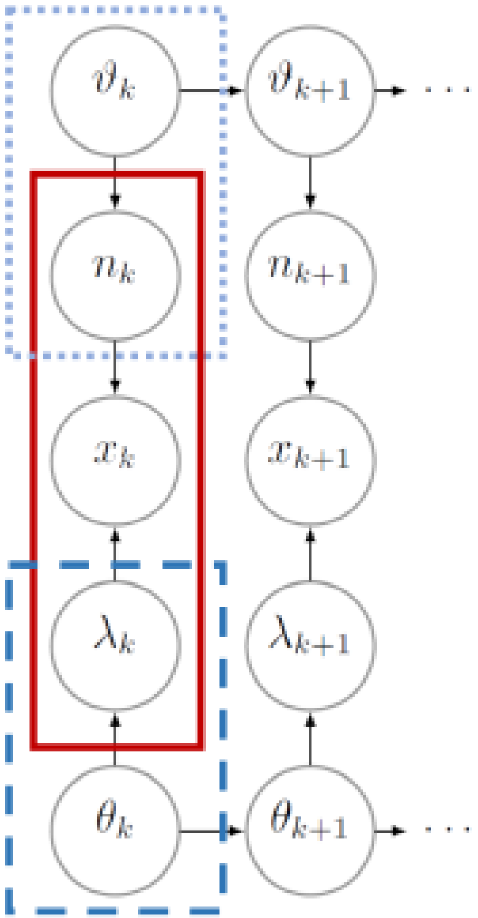
Statistical models as PGMs. Decision Analysis



Statistical models as PGMs. Hierarchical models



National aviation safety plan

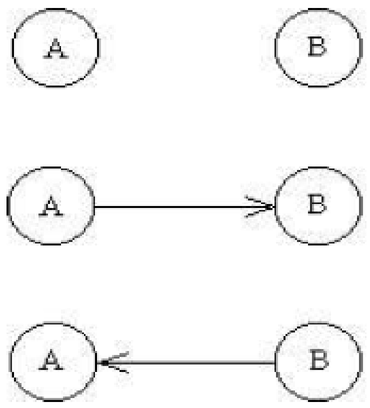


$$
\begin{cases}
\begin{cases}
\begin{cases}
n_k = \boldsymbol{H}_k \vartheta_k + z_k, \quad z_k \sim N(0, \Sigma_k) \\
\vartheta_k = \boldsymbol{J}_k \vartheta_{i-1} + \xi_k, \quad \xi_k \sim N(\boldsymbol{0}, \boldsymbol{S}_k)
\end{cases} \\
\vartheta_0 \sim N(\boldsymbol{\eta}_0, \boldsymbol{S}_0)
\end{cases} \\
x_k | \lambda_k, n_k \sim Po(\lambda_k n_k), \qquad \lambda_k = \exp(u_k) \\
\begin{cases}
\begin{cases}
u_k = \boldsymbol{F}_k \boldsymbol{\theta}_k + v_k, \quad v_k \sim N(0, V_k) \\
\boldsymbol{\theta}_k = \boldsymbol{G}_k \boldsymbol{\theta}_{k-1} + \boldsymbol{w}_k, \quad \boldsymbol{w}_k \sim N(\boldsymbol{0}, \boldsymbol{W}_k)
\end{cases} \\
\boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0),
\end{cases}
\end{cases}
$$

Assessments. Discrete case

1 node

2 nodes



M nodes

$$p(x|\mu) = \prod_{k=1}^{K} \mu_k^{x_k} \longrightarrow K-1$$

$$p(x_1,x_2|\mu) = \prod_{k=1}^{K} \prod_{l=1}^{K} \mu_{kl}^{x_{1k}x_{2l}} \longrightarrow K^2-1$$

$$p(x_1,x_2) = p(x_2|x_1)\,p(x_1) \quad (K-1) + K(K-1) \longrightarrow K^2-1$$

$$p(x_1,x_2) = p(x_1)\,p(x_2) \longrightarrow 2(K-1)$$

FULLY CONNECTED $\longrightarrow K^M-1$

INDEPENDENT $\longrightarrow M(K-1)$

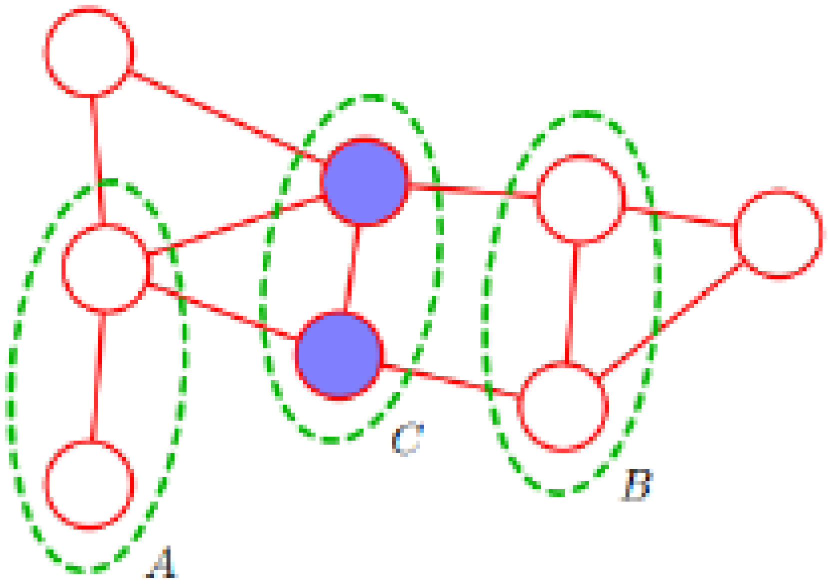
M.CHAIN $(K-1)+(M-1)K(K-1) \longrightarrow O(K^2)$, $O(M)$

Probabilistic graphical models. Undirected Markov random fields

Conditional independence

A , B are conditionally independent given C

If all paths between nodes in A and B are blocked by a node in C



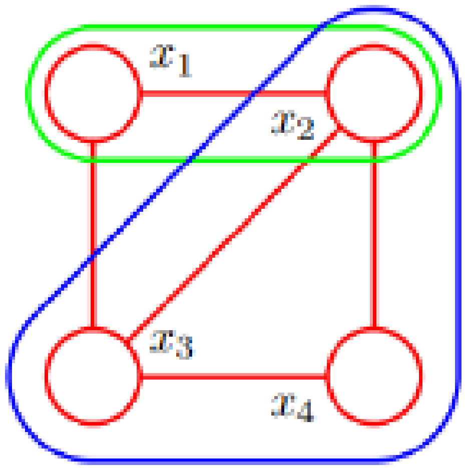
Factorization properties

If nodes x_i and x_j not connected by arc

$$p(x_i, x_j | x_{\backslash\{i,j\}}) = p(x_i | x_{\backslash\{i,j\}}) p(x_j | x_{\backslash\{i,j\}}).$$

Cliques. Subgraphs with every pair of nodes linked by arc

Maximal cliques. No node can be added without losing the clique property



Joint distribution

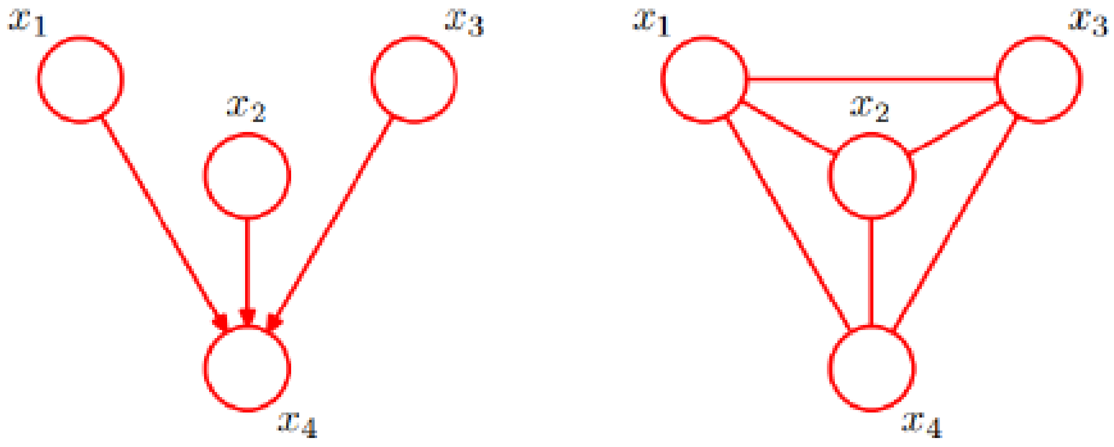
For a clique C and the variables in it we use potentials

$$p(x) = \frac{1}{Z} \prod_C \Psi_C(x_C), \qquad Z = \sum_x \prod_C \Psi_C(x_C)$$

Potential functions do not have probabilistic interpretation, in general

From directed to indirected

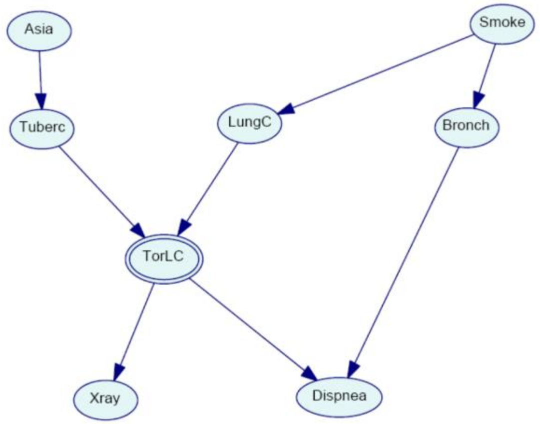
Moralization of graph. Moral graph



Inference in graphical models

General problem

Assuming DAG (arcs and distributions at nodes):

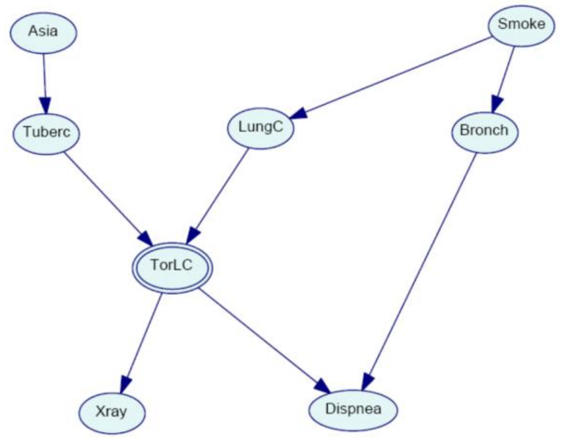


1. Initialisation
2. Absorption of evidence
3. Global propagation of evidence
4. Hypothesising and propagating single pieces of evidence
5. Planning
6. Influential findings

Core ideas

Model $p(\alpha, \tau, \xi, \varepsilon, \delta, \lambda, \beta, \sigma)$

expressed as $p(\alpha)p(\tau \mid \alpha)p(\xi \mid \varepsilon)p(\varepsilon \mid \tau, \lambda)p(\delta \mid \varepsilon, \beta)p(\lambda \mid \sigma)p(\beta \mid \sigma)p(\sigma)$



Typical (probabilistic) query $p(x \mid a, d)$

Trivially $p(x, a, d)/p(a, d)$ and can be computed by brute force…..

Idea 1. Take advantage of structure

$$p(a) \sum_{\tau} p(\tau \mid a)\left[\sum_{\varepsilon} p(x \mid \varepsilon)\left[\sum_{\lambda} p(\varepsilon \mid \tau, \lambda)\left[\sum_{\beta} p(d \mid \varepsilon, \beta)\left[\sum_{\sigma} p(\lambda \mid \sigma)p(\beta \mid \sigma)p(\sigma)\right]\right]\right]\right]$$

Core ideas

Idea 2. Full calculation not needed until the end

$$p(\alpha)p(\tau \mid \alpha)p(\xi \mid \varepsilon)p(\varepsilon \mid \tau, \lambda)p(\delta \mid \varepsilon, \beta)p(\lambda \mid \sigma)p(\beta \mid \sigma)p(\sigma)$$

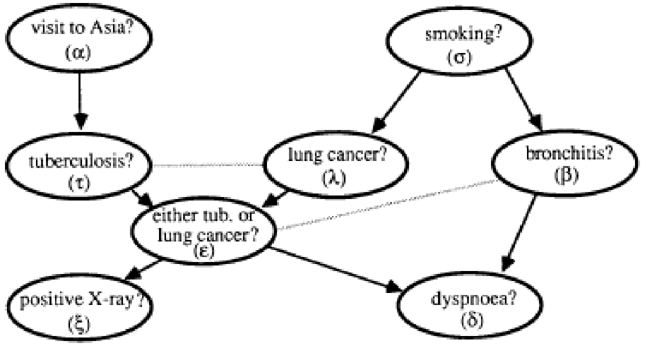
Rewritten (initially) as

$$\psi(\alpha)\psi(\tau, \alpha)\psi(\xi, \varepsilon)\psi(\varepsilon, \tau, \lambda)\psi(\delta, \varepsilon, \beta)\psi(\lambda, \sigma)\psi(\beta, \sigma)\psi(\sigma)$$

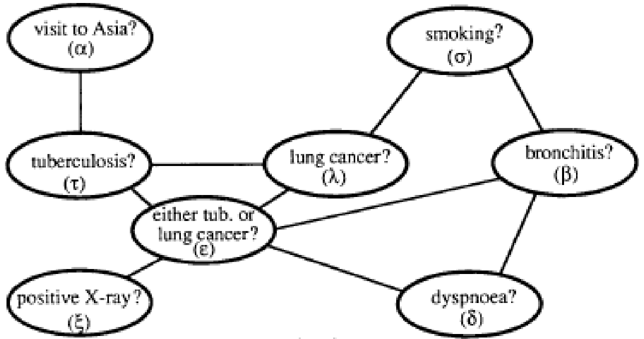
Idea 3. Track computations through moral graph

Idea 4. Actually track it through triangulated mg

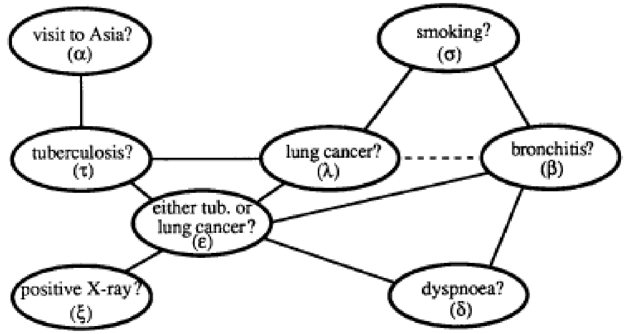
$$p \propto \psi(\alpha, \tau)\psi(\tau, \lambda, \varepsilon)\psi(\lambda, \varepsilon, \beta)\psi(\lambda, \beta, \sigma)\psi(\varepsilon, \beta, \delta)\psi(\varepsilon, \xi$$



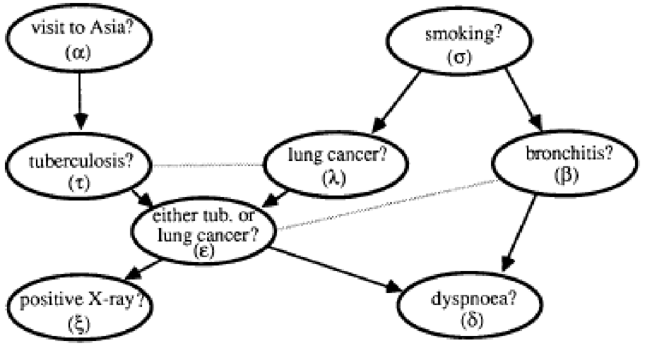
marrying parents



dropping directions

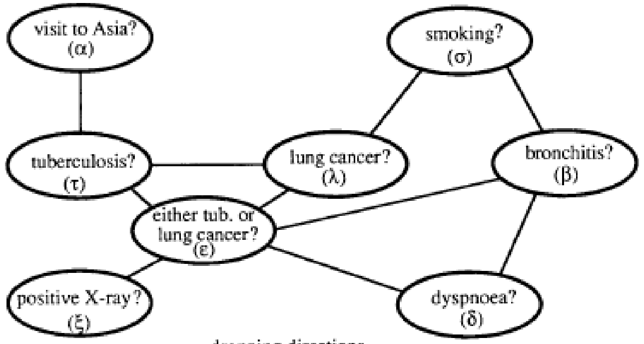


Core ideas



marrying parents

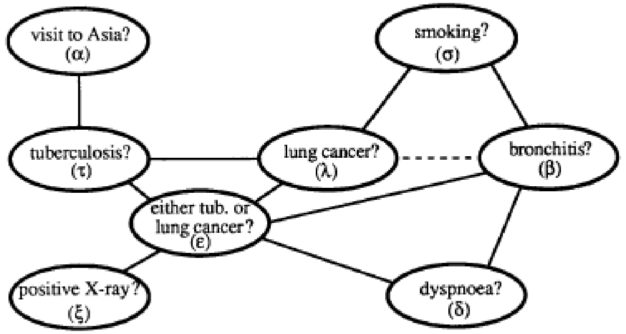
Idea 6. Represent joint in terms of marginals on cliques



dropping directions

$$\frac{p(\alpha, \tau)p(\tau, \lambda, \varepsilon)p(\lambda, \varepsilon, \beta)p(\lambda, \beta, \sigma)p(\varepsilon, \beta, \delta)p(\varepsilon, \xi)}{p(\tau)p(\lambda, \varepsilon)p(\lambda, \beta)p(\varepsilon, \beta)p(\varepsilon)}$$

Idea 7. Store clique marginals



Algos

Sum-product

Max-product

Junction tree

……

Simulation based

Sampling from a belief network. Generative model

$$p(\mathbf{x}) = \prod_i p\left(\mathrm{x}_i \mid Pa_{\mathcal{G}}(\mathrm{x}_i)\right)$$

For $i=1$ to n

Sample $x_i \sim p(x_i \mid Pa_{\mathcal{G}}(x_i))$

Generic Gibbs sampler

$$X = (X_1, \dots, X_p) \sim \pi$$

Sample from $X_s \mid X_{-s} = (X_1, \dots, X_{s-1}, X_{s+1}, \dots, X_p)$

Initialize $X_1^0, \dots, X_p^0$, $i = 1$

Iterate

Sample $X_1^i \sim X_1 \mid X_2^{i-1}, \dots, X_p^{i-1}$

Sample $X_2^i \sim X_2 \mid X_1^i, X_3^{i-1}, \dots, X_p^{i-1}$

...

Sample $X_p^i \sim X_p \mid X_1^i, X_2^i, \dots, X_p^{i-1}$

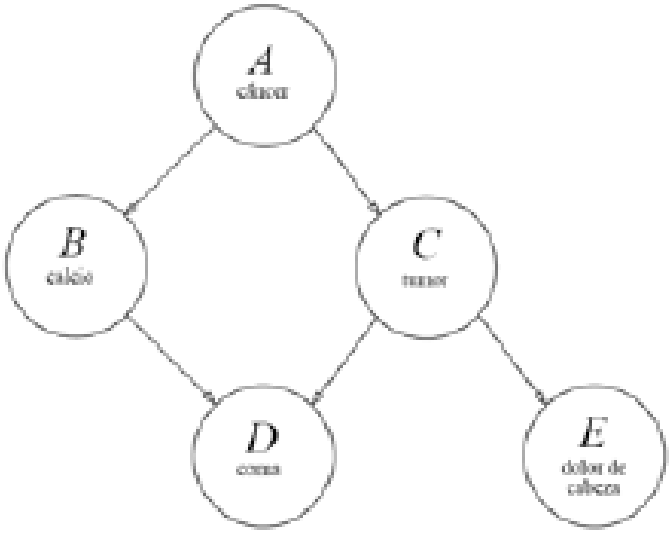
$i = i+1$

Gibbs sampler for belief nets

Conditionals

$$P(X_j = x_j | X_{-j} = x_{-j}) = \alpha P(X_j = x_j | \Pi_{X_j}(x_{-j})) \prod_{Y_j \in S_j} P(Y_j = y_j | \Pi_{Y_j}(x_j))$$

Back to example



a	0.2

	a	$\overline{a}$
b	0.8	0.2

	a	$\overline{a}$
c	0.2	0.05

	b, c	$\overline{b}, c$	$b, \overline{c}$	$\overline{b}, \overline{c}$
d	0.8	0.8	0.8	0.05

	c	$\overline{c}$
e	0.8	0.6

$$P(c|\overline{d}, e) = \frac{P(c, \overline{d}, e)}{P(\overline{d}, e)}$$

$$\begin{aligned} P(c, \overline{d}, e) &= \sum_{\alpha,\beta} P(\alpha, \beta, c, \overline{d}, e) = \sum_{\alpha,\beta} P(\alpha)P(\beta|\alpha)P(c|\alpha)P(\overline{d}|\beta, c)P(e|c) \\ &= P(a)P(b|a)P(c|a)P(\overline{d}|b, c)P(e|c) + P(a)P(\overline{b}|a)P(c|a)P(\overline{d}|\overline{b}, c)P(e|c)+ \\ &\quad P(\overline{a})P(b|\overline{a})P(c|\overline{a})P(\overline{d}|b, c)P(e|c) + P(\overline{a})P(\overline{b}|\overline{a})P(c|\overline{a})P(\overline{d}|\overline{b}, c)P(e|c) \\ &= 0.0118 \end{aligned}$$

$$P(\overline{d}, e) = \sum_{\alpha,\beta,\gamma} P(\alpha, \beta, \gamma, \overline{d}, e) = 0.410$$

$$P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)P(E|C)$$

$$P(c|\overline{d}, e) = 0.0287$$

Back to example

$$
\begin{aligned}
P(A|B,C,\bar{d},e) &= P(A|x_{-A}) = \alpha_1 P(A)P(B|A)P(C|A)\\
P(B|A,C,\bar{d},e) &= P(B|x_{-B}) = \alpha_2 P(B|A)P(\bar{d}|B,C)\\
P(C|A,B,\bar{d},e) &= P(C|x_{-C}) = \alpha_3 P(C|A)P(\bar{d}|B,C)P(e|C)
\end{aligned}
$$

```
Seleccionar B = b_0, C = c_0 arbitrariamente
Hacer j = 1
Mientras no se juzgue convergencia,
    Generar A_j = a_j ~ P(A|x_{-A}) = α_{1j} P(A) P(b_{j-1}|A) P(c_{j-1}|A)
    Generar B_j = b_j ~ P(B|x_{-B}) = α_{2j} P(B|a_j) P(d̄|B, c_{j-1})
    Generar C_j = c_j ~ P(C|x_{-C}) = α_{3j} P(C|a_j) P(d̄|b_j, C) P(e|C)
    Hacer j = j + 1
```

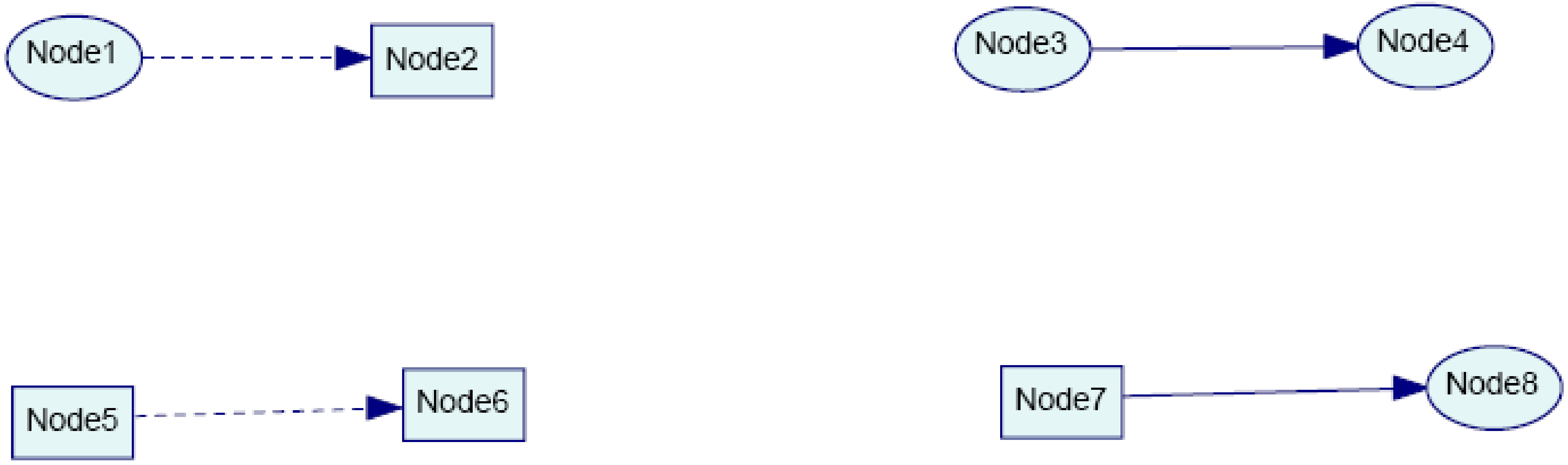
$$\frac{\#\{C_j = c\}}{M}$$

Final comments: Influence diagrams

Influence Diagrams

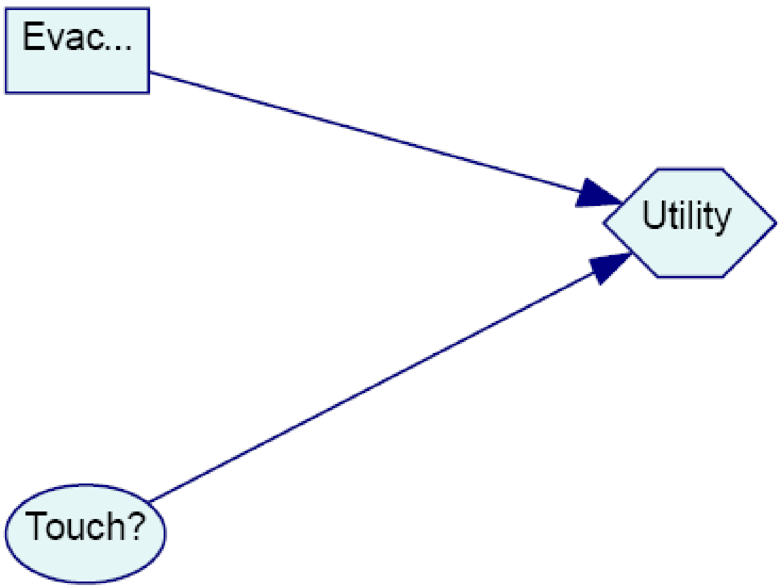
- Tool to structure (and solve) decision making problems
- Graph with nodes and arcs. No cycles
- Three main types of nodes.
  - Chance. Circle
  - Decision. Square
  - Value. Hexagon, Diamond
    - Fourth type of node. Deterministic. Double circle
- Two types of arcs
  - Arcs into decision nodes
  - Arcs into chance and value nodes

Influence Diagrams. Interpretation?



Suppose you’re Nags Head mayor. There is a hurricane threat.
Would you issue an evacuation order?

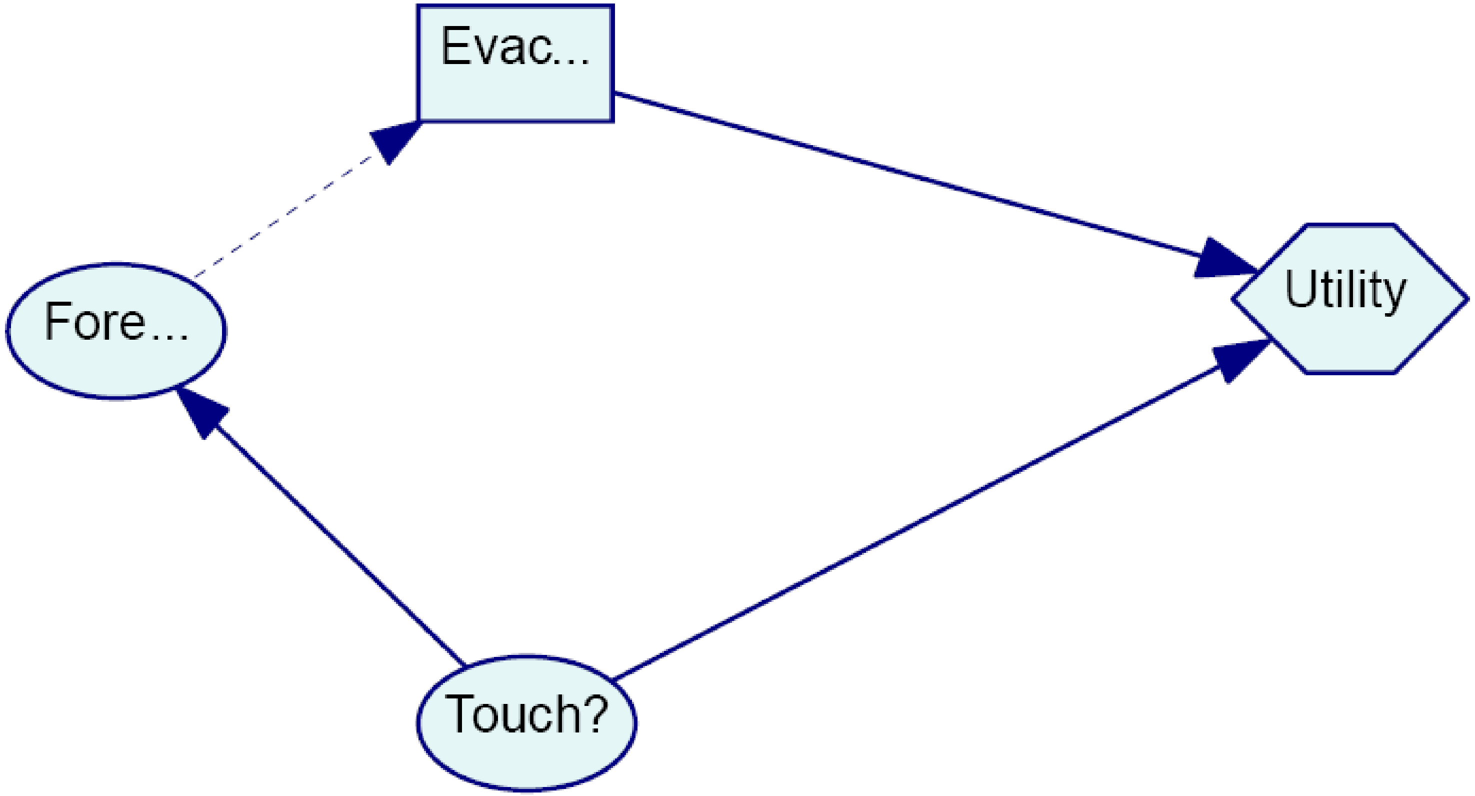
Decision under risk



Suppose you’re Nags Head mayor. There is a hurricane threat. Would you issue an evacuation order?

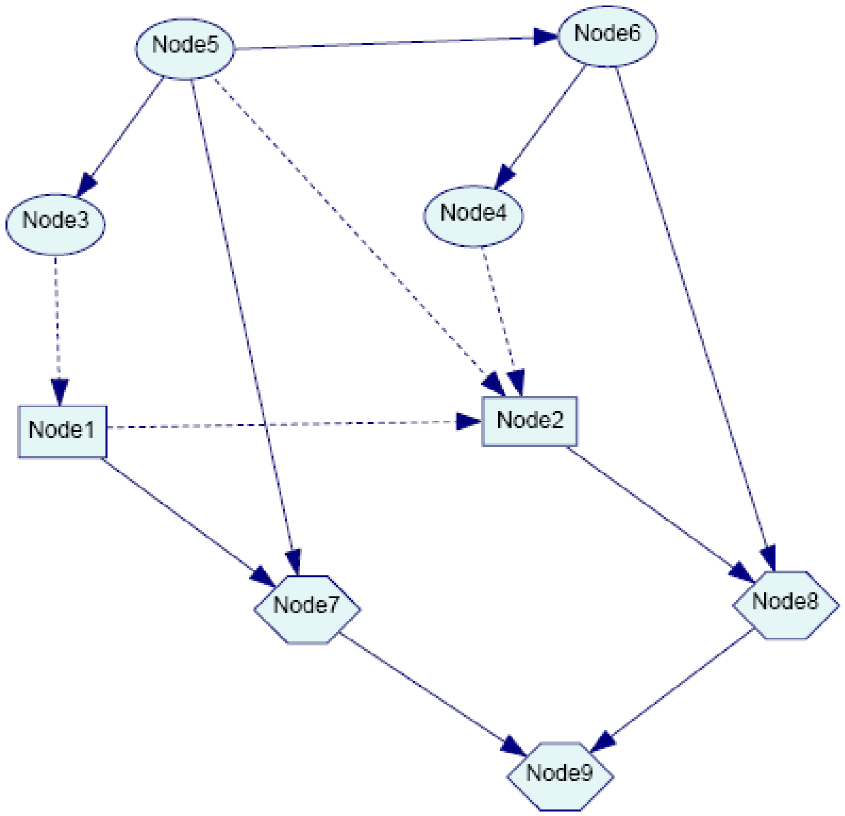
You have as info a forecast from the NHC. But the forecast is not perfect…

Decision under risk with imperfect information



… the problem repeated over the hurricane season….

Sequential Decisions.



Additional comments

Learning structure from data: **Structure learning**. Greedy search based on a scoring function based on an information measure

Learning node distributions….

Back when deep learning. Deep belief nets

GeNIe

- https://www.bayesfusion.com/influence-diagrams/
- https://download.bayesfusion.com/files.html?category=Academia

Back when reinforcement learning

See you next week

introml@icmat.es

Stuff at

https://datalab-icmat.github.io/courses_stats.html