Propagation in Trees

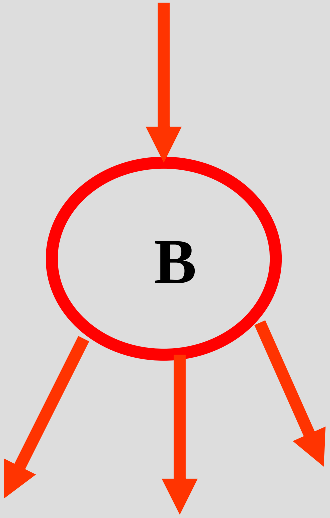
- Message passing algorithm

- Each node is a discrete random variable, B, with values: {B1,B2,…,Bn}; and corresponding conditional probability matrix: P(B|A)=P(Bj| Ai)

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Given certain evidence *E* --instantiation of some variables--, the posterior probability of any variable *B*, by Bayes theorem:

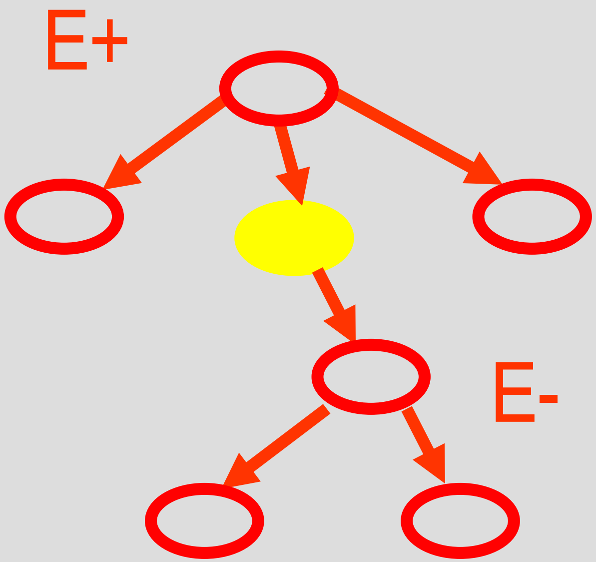
$$P(B_i \mid E) = P(B_i)\, P(E \mid B_i) / P(E)$$



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Propagation in trees

- Given that the BN has a tree structure, any node (B) separates the evidence in two parts:
- E-: Data in the tree which root is B
- E+: Data in the rest of the network



Propagation in trees

- Then:

$$P(B_i \mid E) = P(B_i)\, P(E^-, E^+ \mid B_i) / P(E)$$

- Given that E^-, E^+ are conditionally independent given B:

$$P(B_i \mid E) = \alpha\, P(B_i \mid E^+)\, P(E^- \mid B_i)$$

Where α is a normalizing constant

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Definitions:

$$\lambda(B_i) = P(E^- \mid B_i)$$

$$\pi(B_i) = P(B_i \mid E^+)$$

Then:

$$P(B_i \mid E) = \alpha\, \pi(B_i)\, \lambda(B_i)$$

Algorithm

- Each node stores the vectors, π and λ, and the conditional probability matrix P

- Probability propagation is done through a message passing mechanism in which each node sends messages to its parents and sons

Message to parent (upwards) -- node B to A:

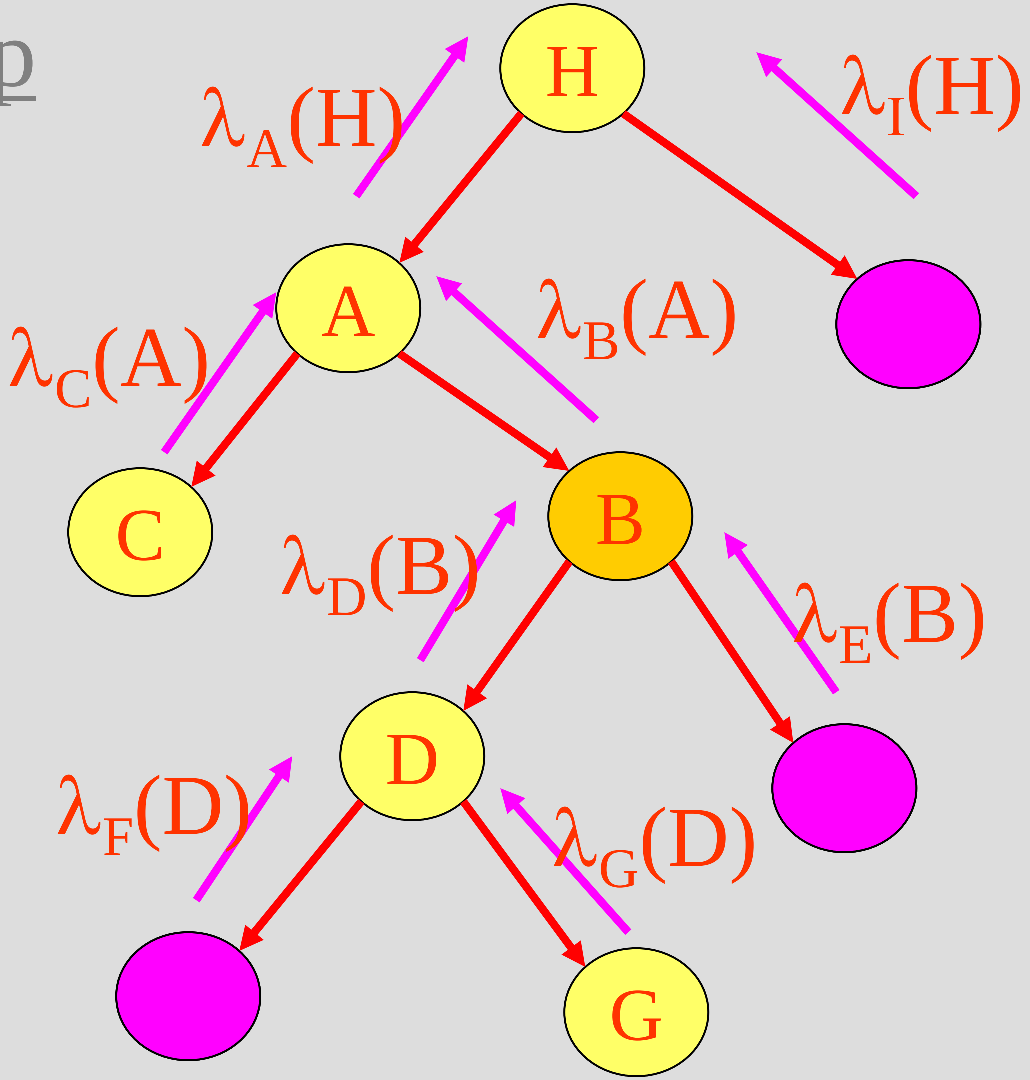
$$\lambda_B(A_i) = \sum_j P(B_j|A_i)\lambda(B_j)$$

Message to sons (downwards) -- node B to son S_k :

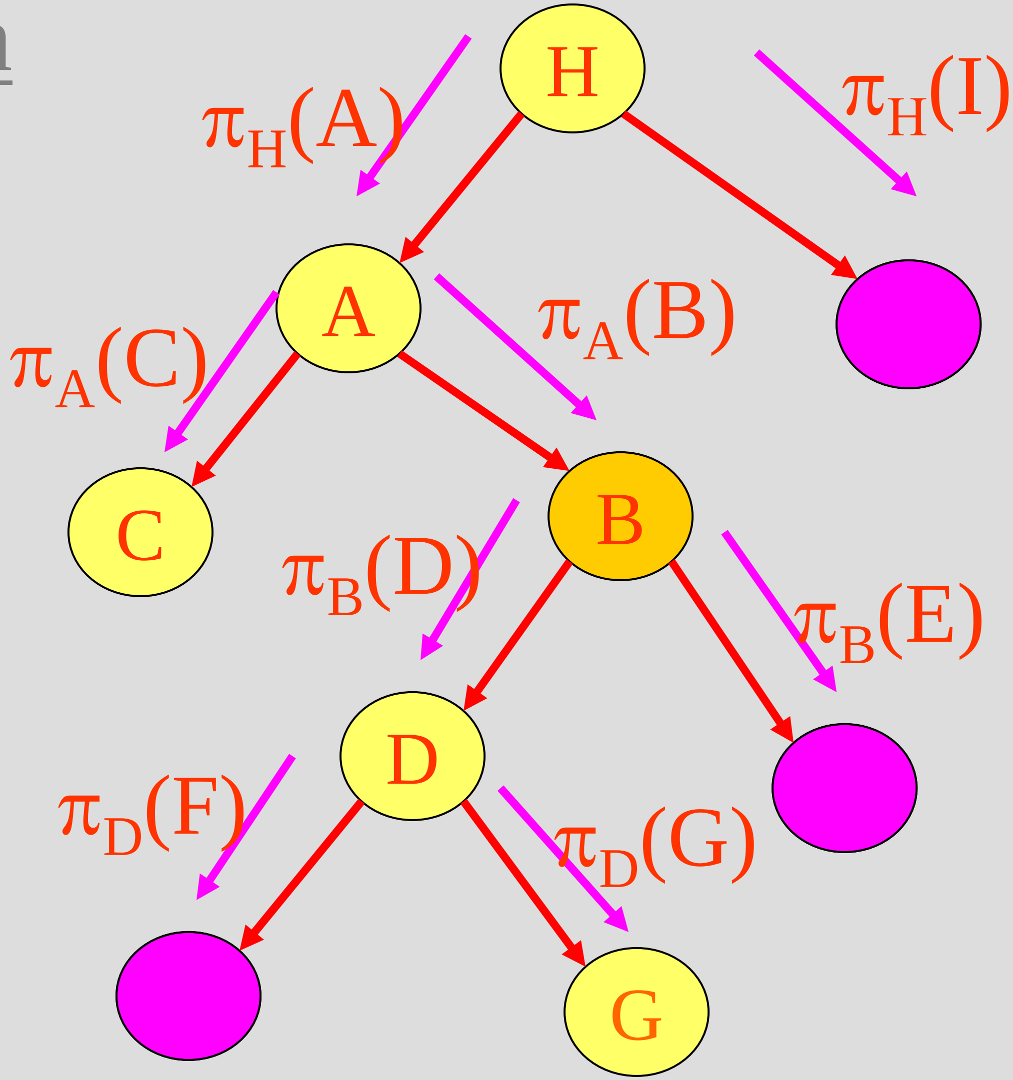
$$\pi_k(B_i) = \alpha\pi(B_j)\prod_{l \neq k} \lambda_l(B_j)$$

Olivier.Aycard@imag.fr
esucar@inaoep.mx
INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA
UNIVERSITE JOSEPH FOURIER

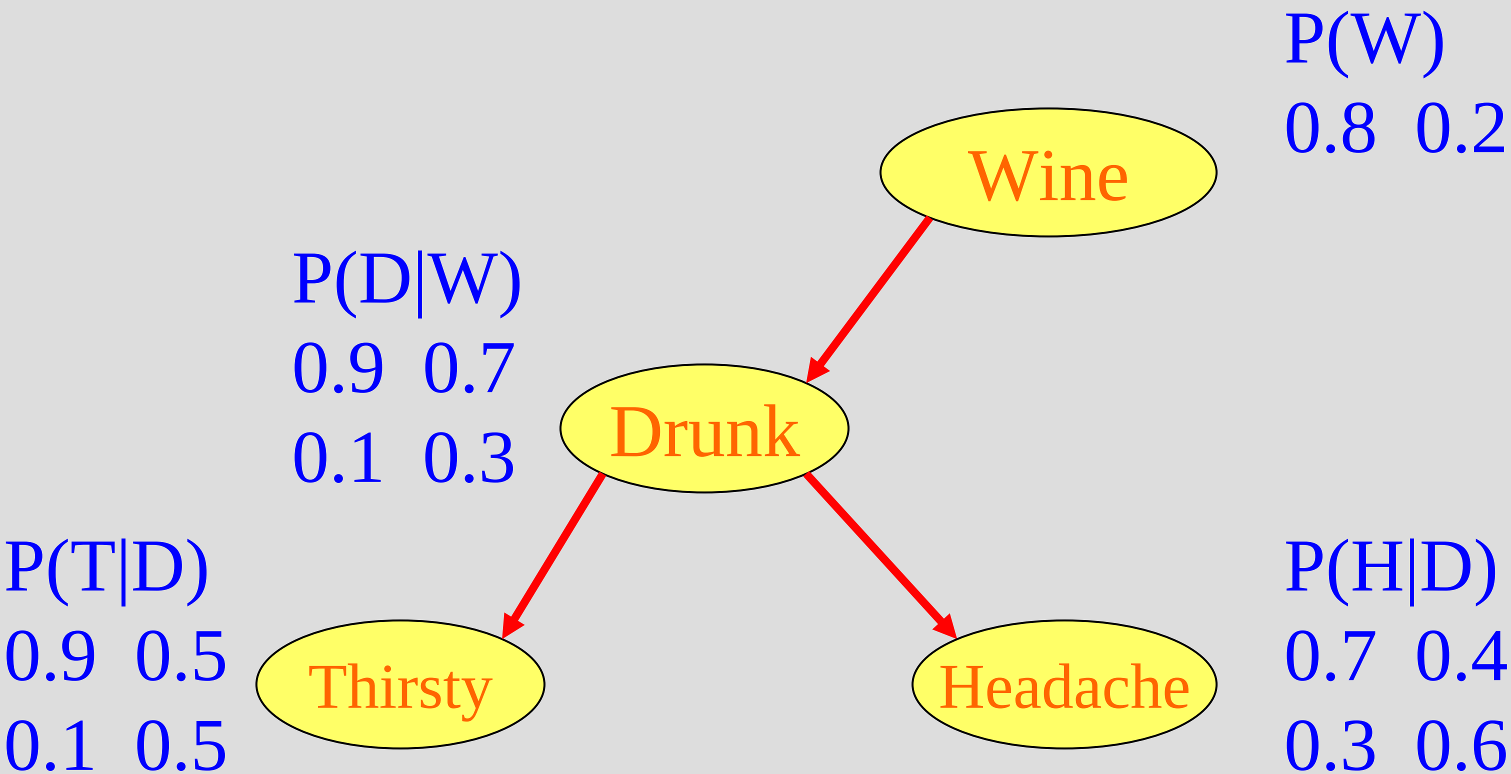
Bottom-up (λ)



Top-down (π)

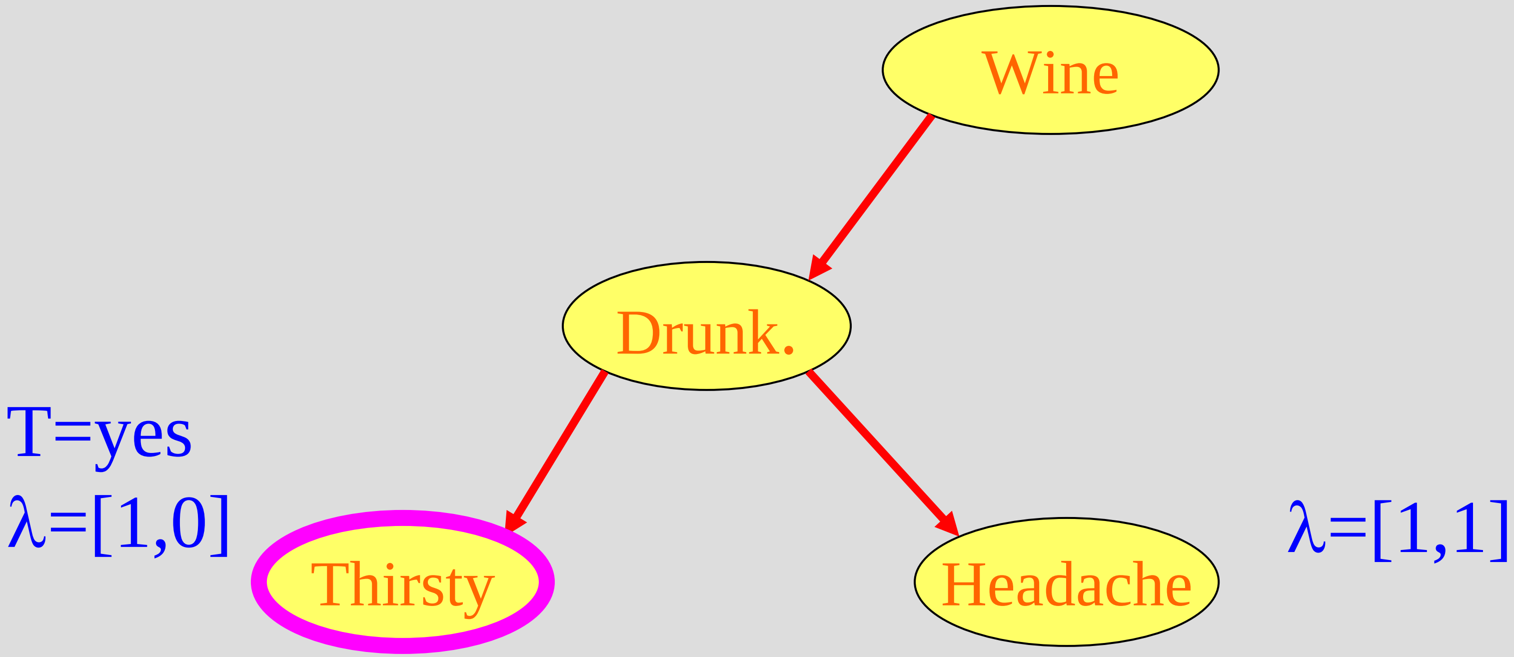


Example



Olivier.Aycard@imag.fr
esucar@inaoep.mx

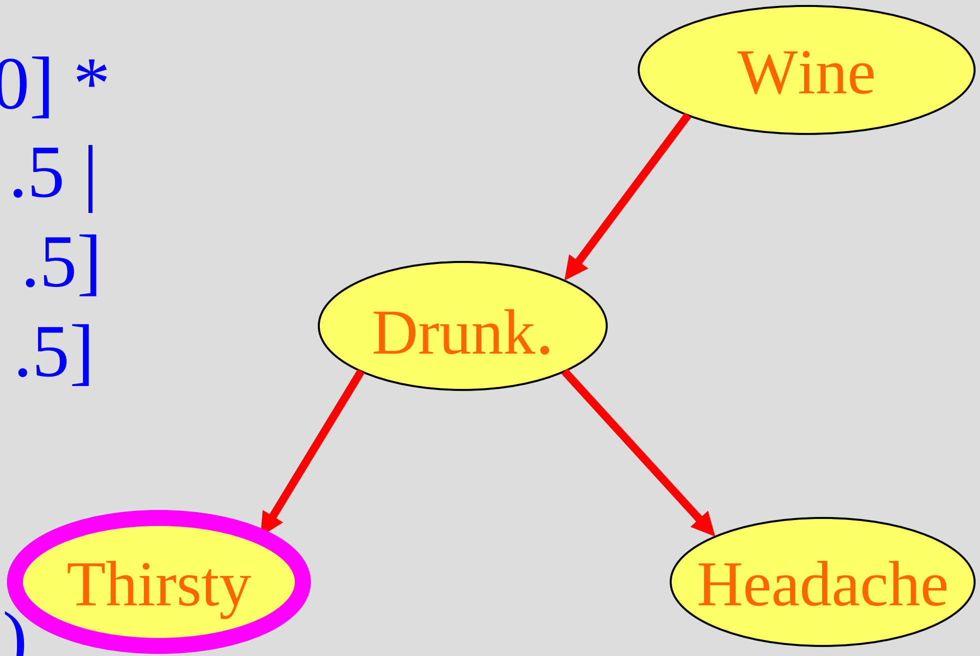
Example



Olivier.Aycard@imag.fr
esucar@inaoep.mx
INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA
UNIVERSITE JOSEPH FOURIER

Example

λ_T= [1,0] *
[.9 .5 |
.1 .5]
= [.9 .5]



λ_H= [1,1] *
[.7 .4 |
.3 .6]
= [1 1]

P(T|D)
0.9 0.5
0.1 0.5

P(H|D)
0.7 0.4
0.3 0.6

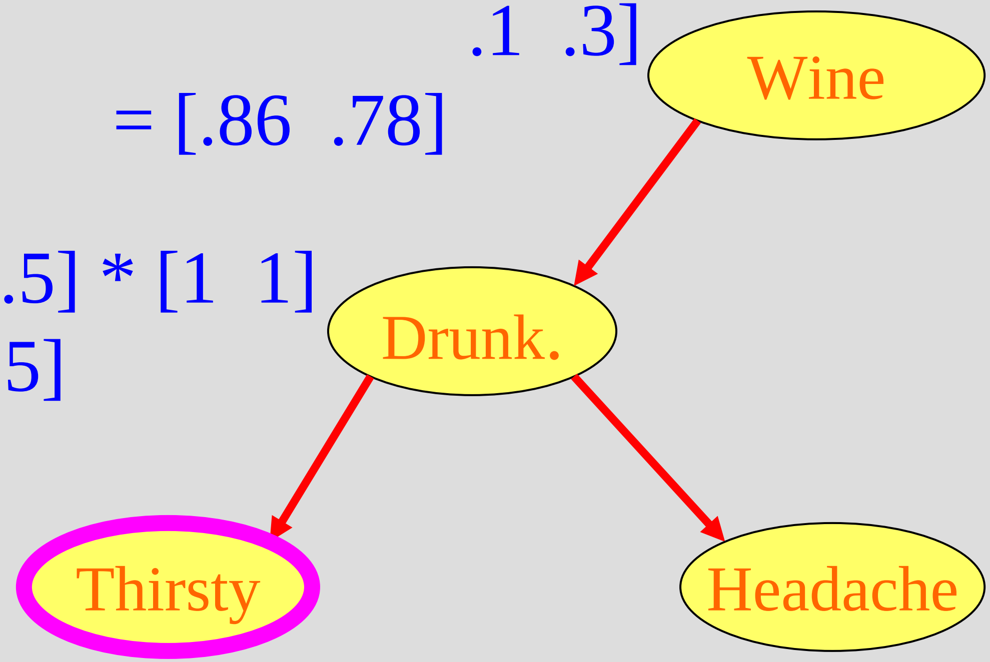
Example

$\lambda(W) = [.9\ .5] * \begin{bmatrix} .9 & .7 \\ .1 & .3 \end{bmatrix}$

$= [.86\ \ .78]$

$\lambda(D) = [.9\ .5] * [1\ \ 1]$

$= [.9\ .5]$



P(D|W)

0.9	0.7
0.1	0.3

P(T|D)

0.9	0.5
0.1	0.5

P(H|D)

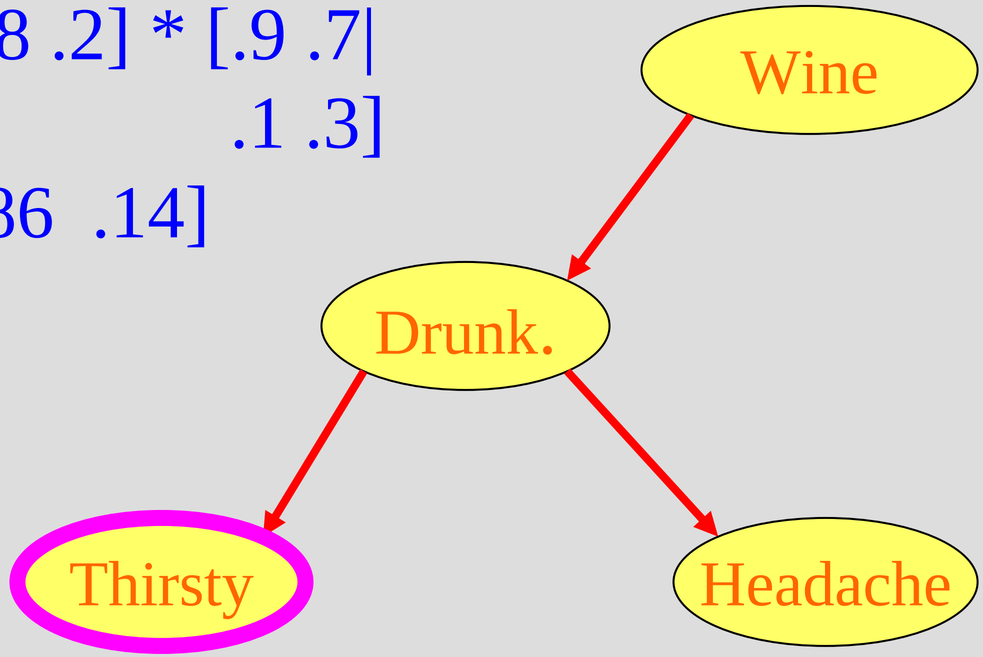
0.7	0.4
0.3	0.6

Olivier.Aycard@imag.fr
esucar@inaoep.mx
INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA
UNIVERSITE JOSEPH FOURIER

Example

$\pi(W) = [.8 \;\; .2]$

$$\pi(D) = [.8 \;\; .2] * \begin{bmatrix} .9 & .7 \\ .1 & .3 \end{bmatrix} = [.86 \;\; .14]$$



P(D|W)
0.9 0.7
0.1 0.3

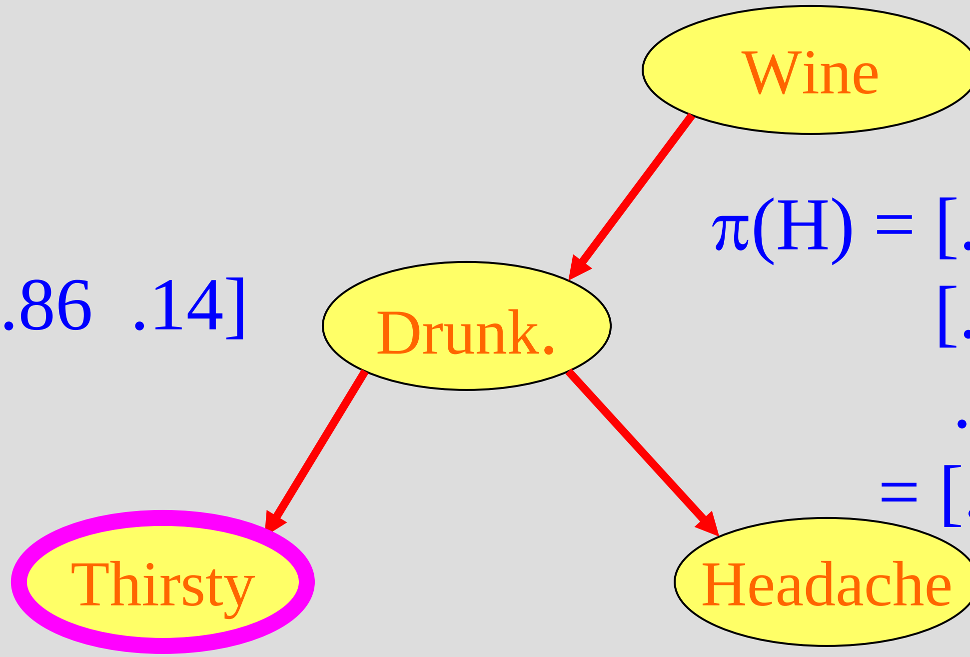
P(T|D)
0.9 0.5
0.1 0.5

P(H|D)
0.7 0.4
0.3 0.6

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Example

$\pi(W) = [.8 \;\; .2]$



$\pi(D) = [.86 \;\; .14]$

$\pi(H) = [.86 \; .14] * [.9 \;\; .5]$

$[.7 \; .4|$

$.3 \; .6]$

$= [.5698 \;\; .2742]$

P(H|D)

0.7 0.4

0.3 0.6

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Example

$\pi(W) = [.8 \;\; .2]$

$\lambda(W) = [.86 \;\; .78]$

$P(W) = \alpha[.688 \;\; .156]$

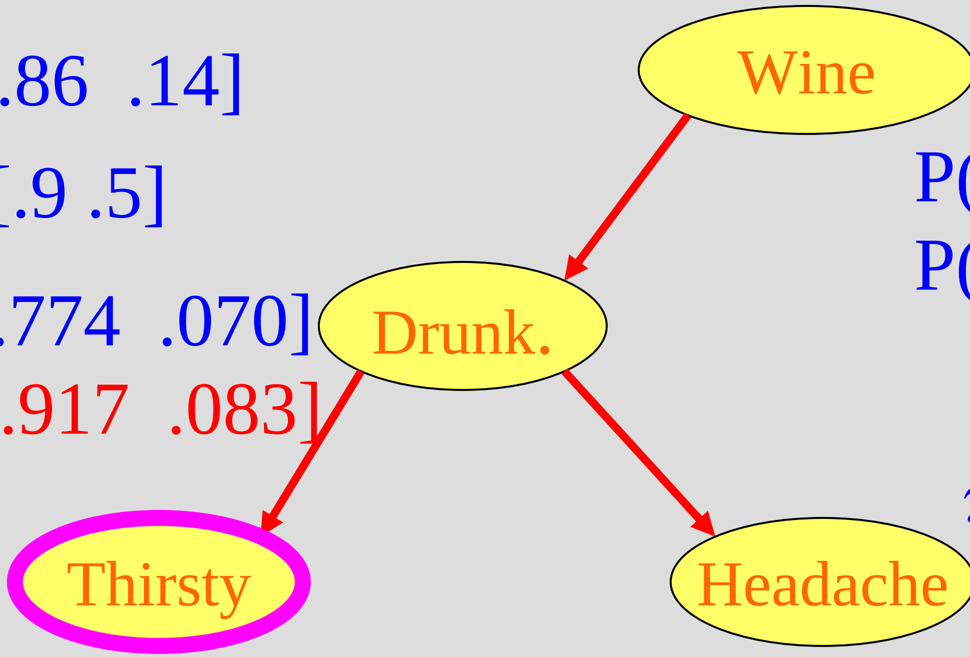
$P(W) = [.815 \;\; .185]$

$\pi(D) = [.86 \;\; .14]$

$\lambda(D) = [.9 \; .5]$

$P(D) = \alpha[.774 \;\; .070]$

$P(D) = [.917 \;\; .083]$



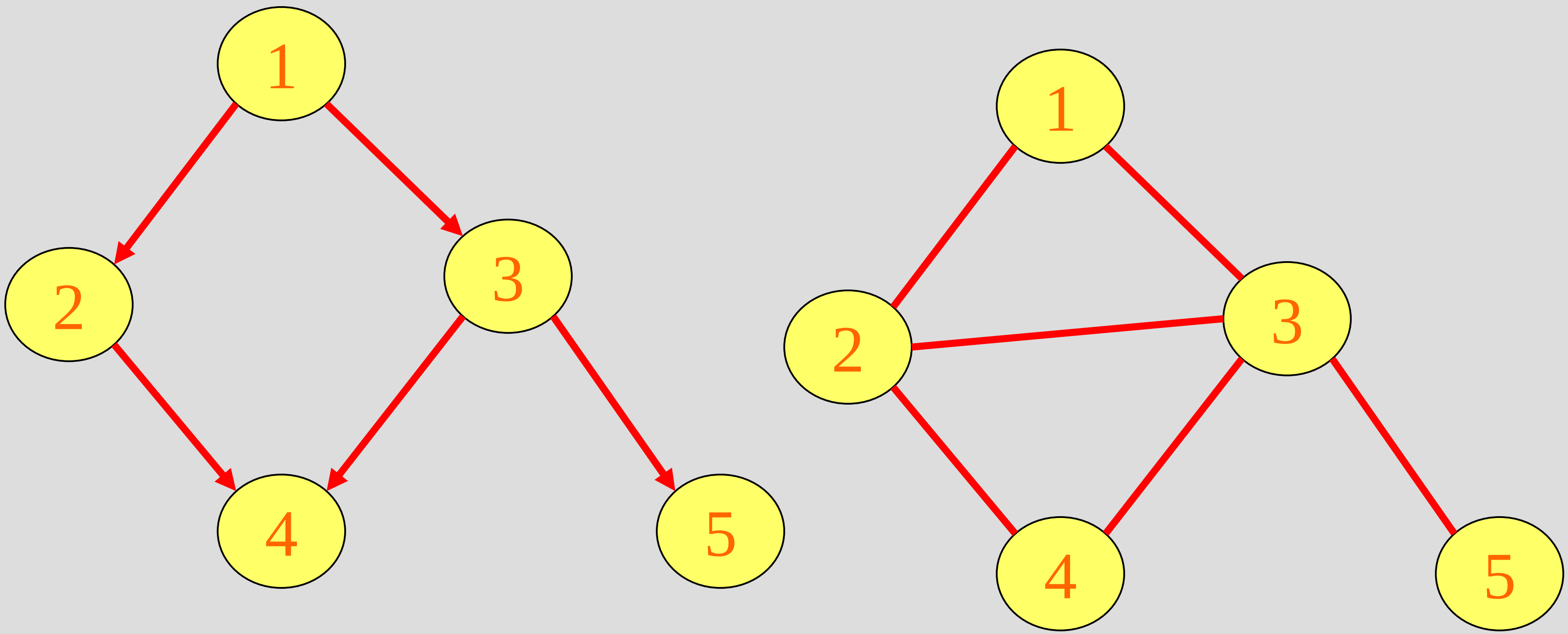
$\pi(H) = [.57 \;\; .27]$

$\lambda(H) = [1,1]$

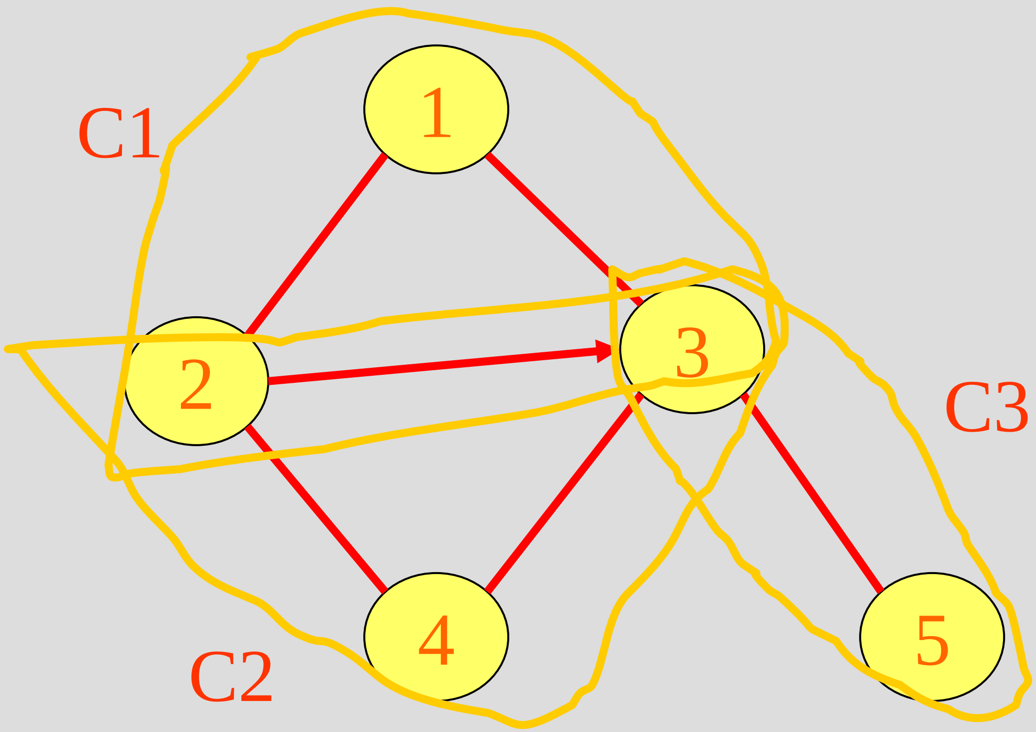
$P(H) = \alpha[.57 \;\; .27]$

$P(H) = [.67 \;\; .33]$

Inference in multiconnected networks: Junction Tree

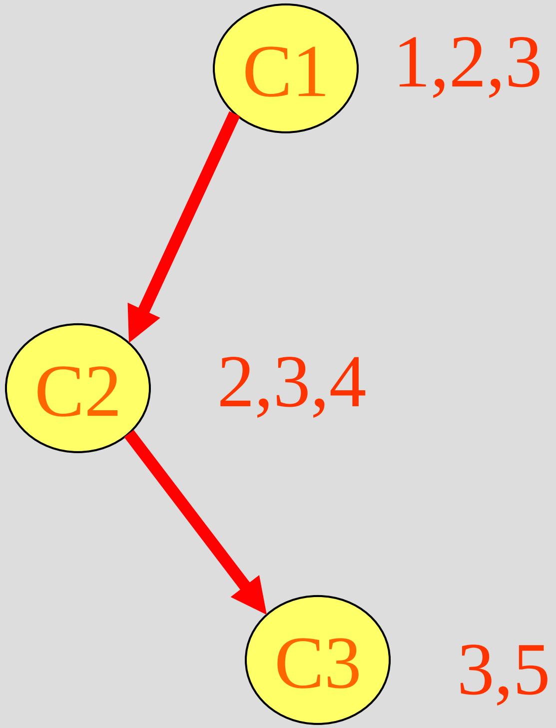


Junction Tree

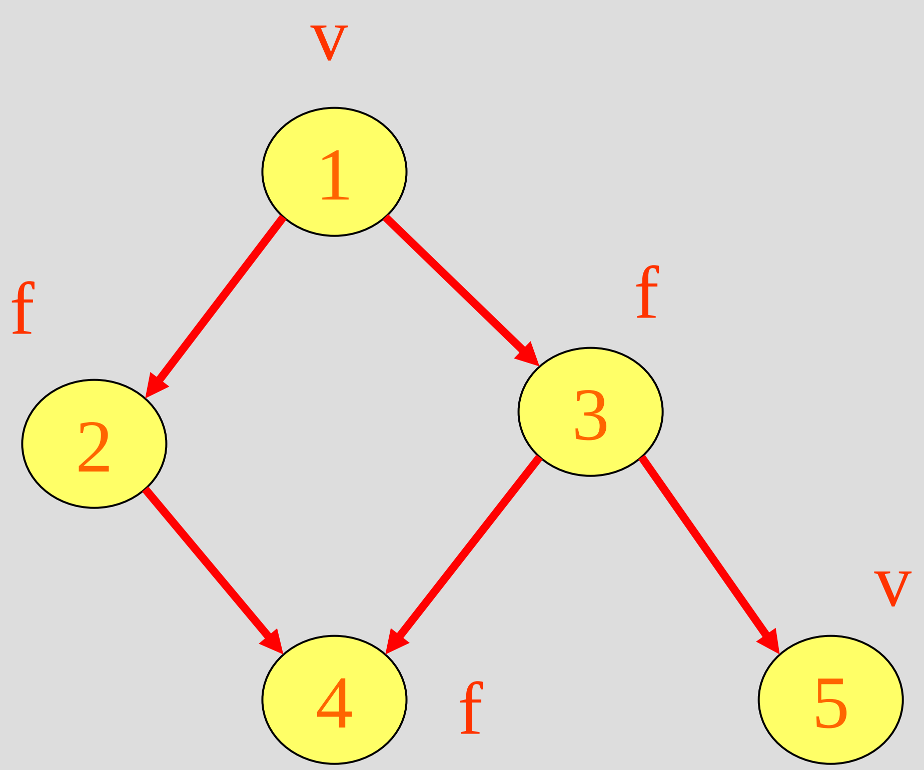


Olivier.Aycard@imag.fr
esucar@inaoep.mx

Junction Tree



Stochastic Simulation

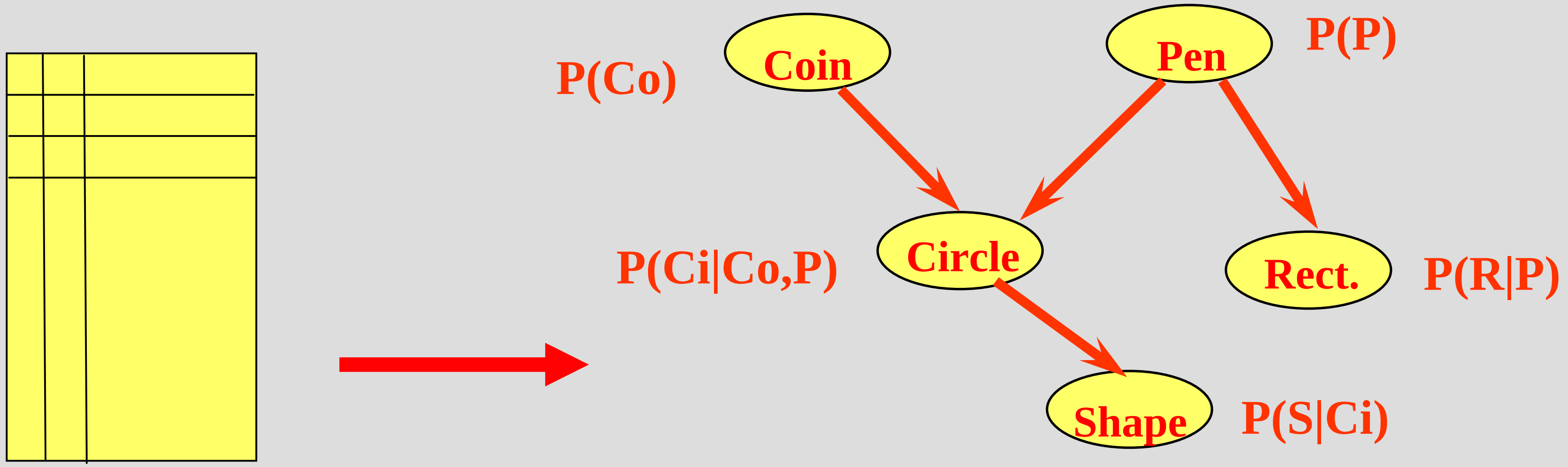


vfffv

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Learning

- Learning in Bayesian networks can be divided into two aspects:
  - Structure Learning
  - Parameter Learning

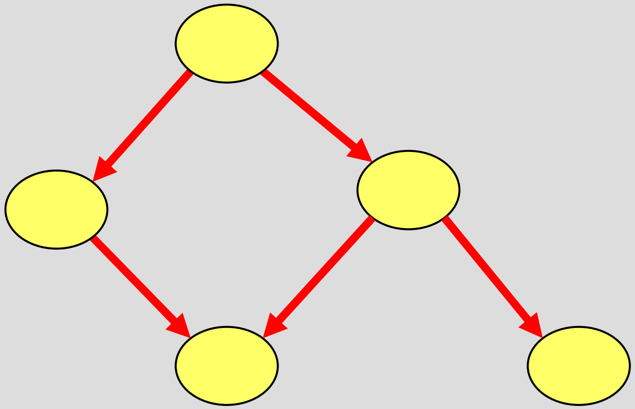


Olivier.Aycard@imag.fr
esucar@inaoep.mx

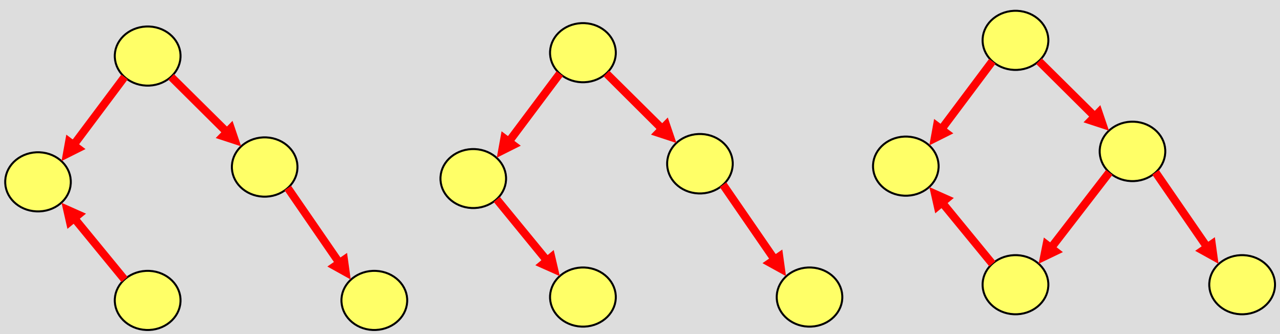
Structure Learning

Two general schemes:

Independence tests



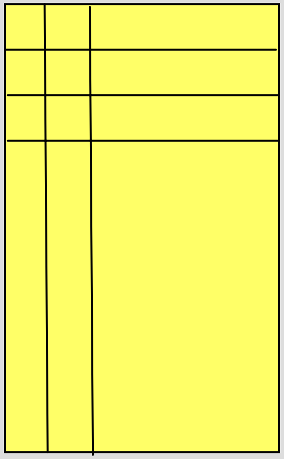
Search and score



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Parameter Learning

- Parameter estimation
  - Maximum likelihood
  - Bayesian (prior)

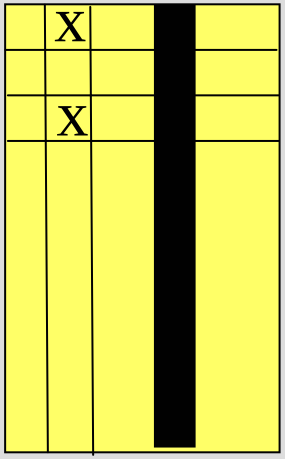


$P(A) = 9/14$

$$P(A) = \frac{(5+9)}{(10+14)}$$

Problems:

- Incomplete data
- Not enough data
- Hidden variables



Statistical Techniques (EM)

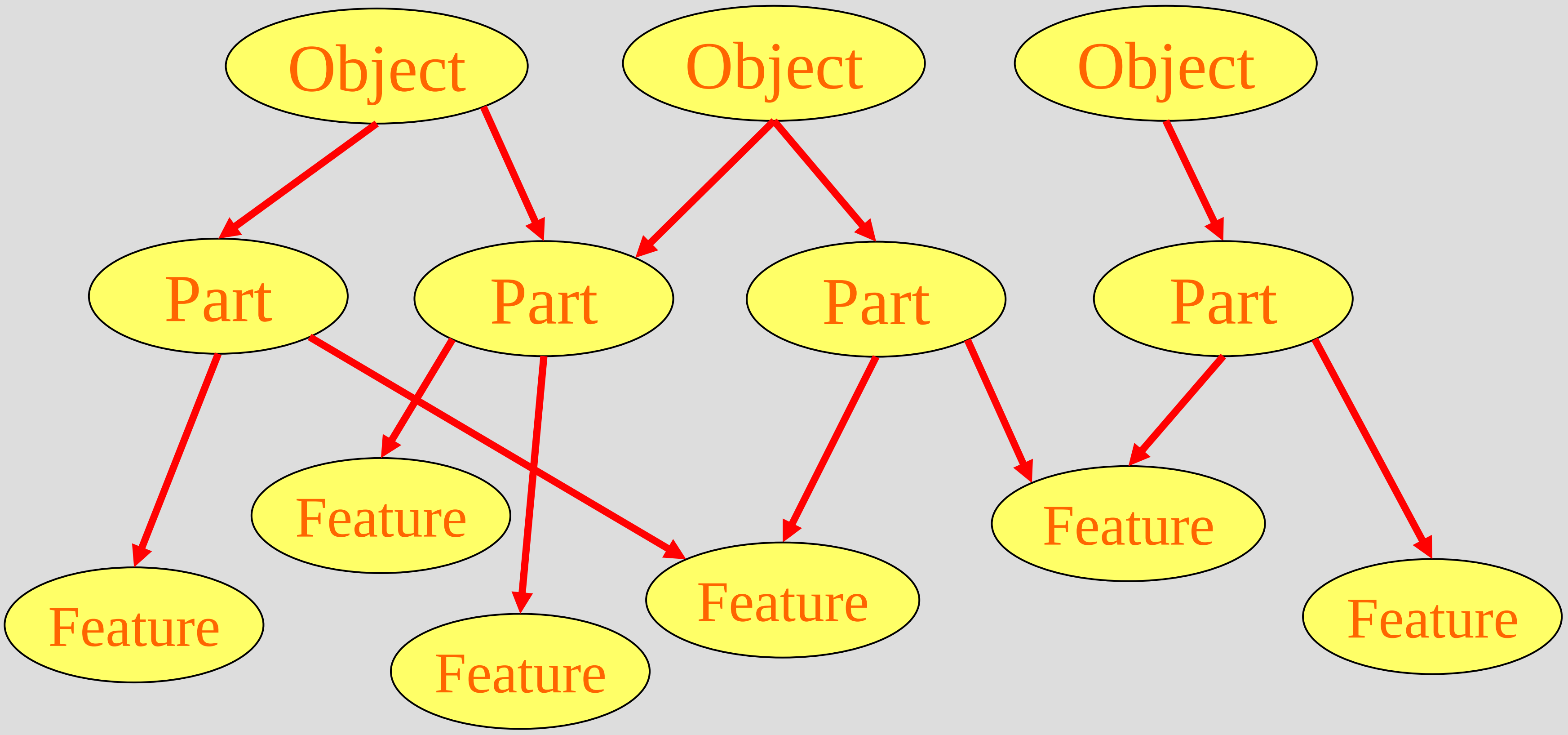
Olivier.Aycard@imag.fr
esucar@inaoep.mx

BN in Vision

General Model

Example

A "general" BN model for Vision

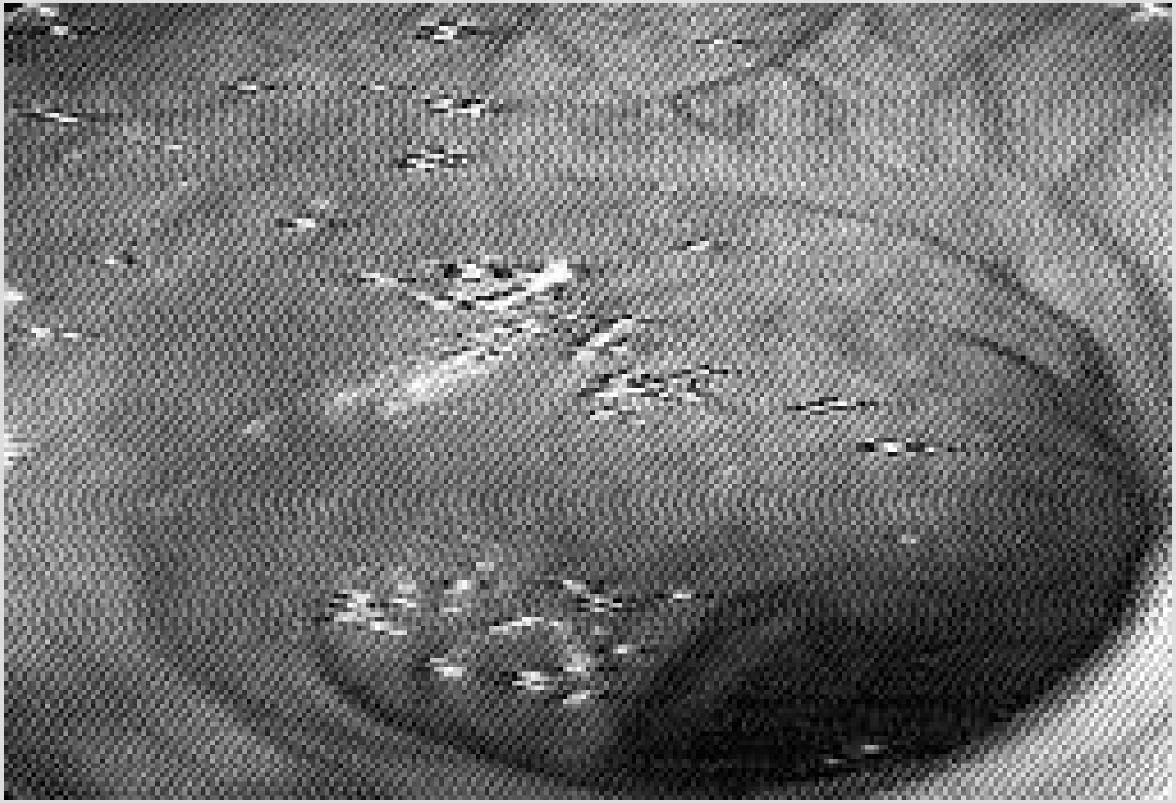


Olivier.Aycard@imag.fr
esucar@inaoep.mx
INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA
UNIVERSITE JOSEPH FOURIER

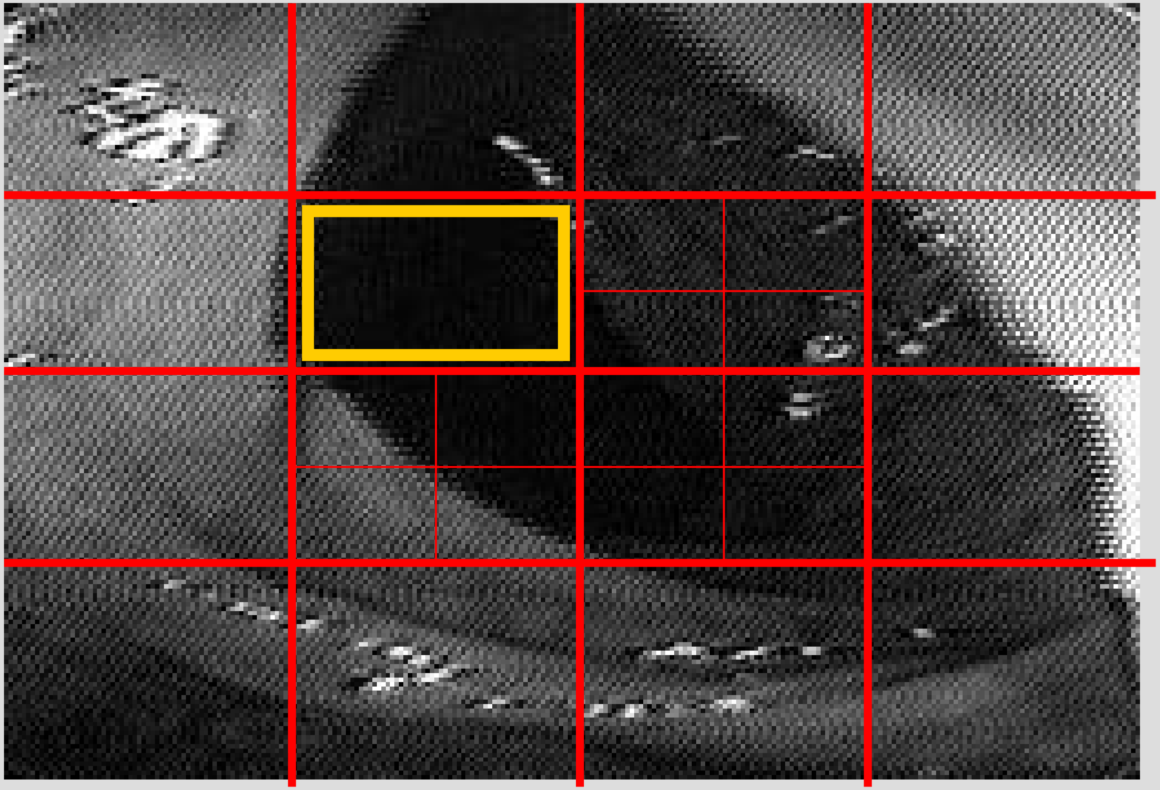
Example - endoscopy

- Endoscopy is a tool for direct observation of the human digestive system
- Recognize "objects" in endoscopy images of the colon for semi-automatic navigation
- Main feature – dark regions
- Main objects – "*lumen*" & "*diverticula*"

Colon Image



Segmentation – dark region

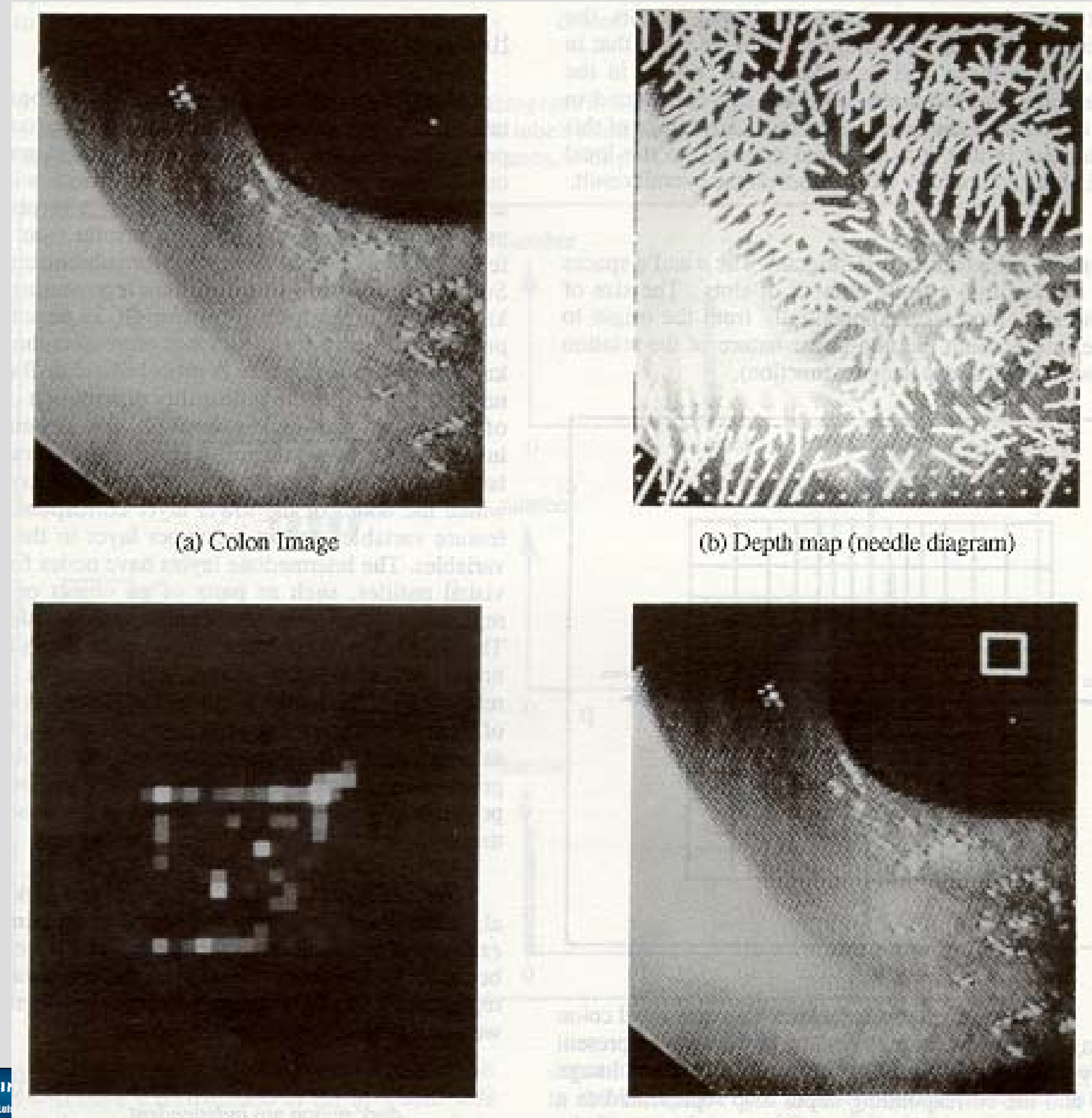


Olivier.Aycard@imag.fr
esucar@inaoep.mx

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

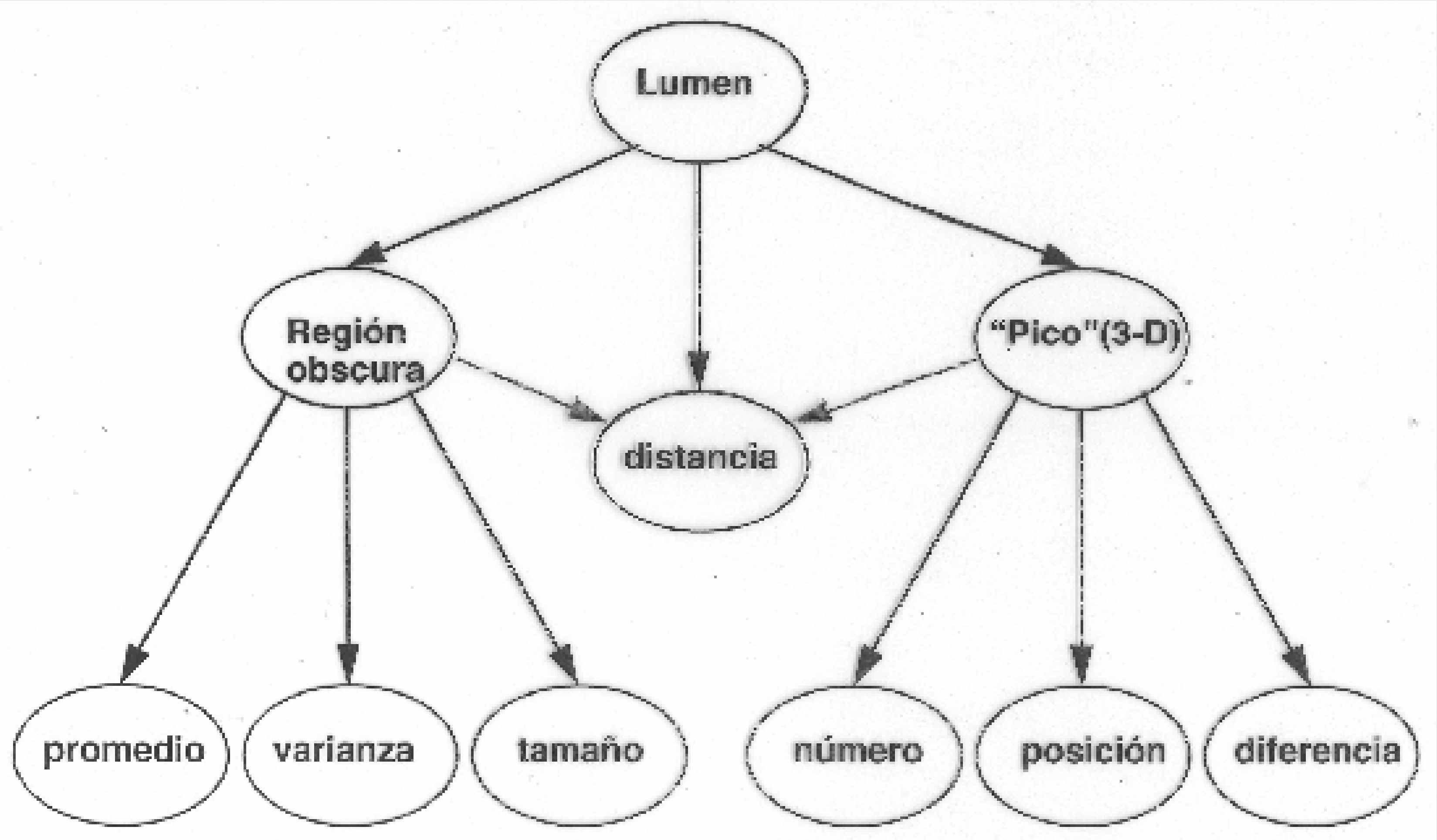
UNIVERSITE JOSEPH FOURIER

Features – pq histogram

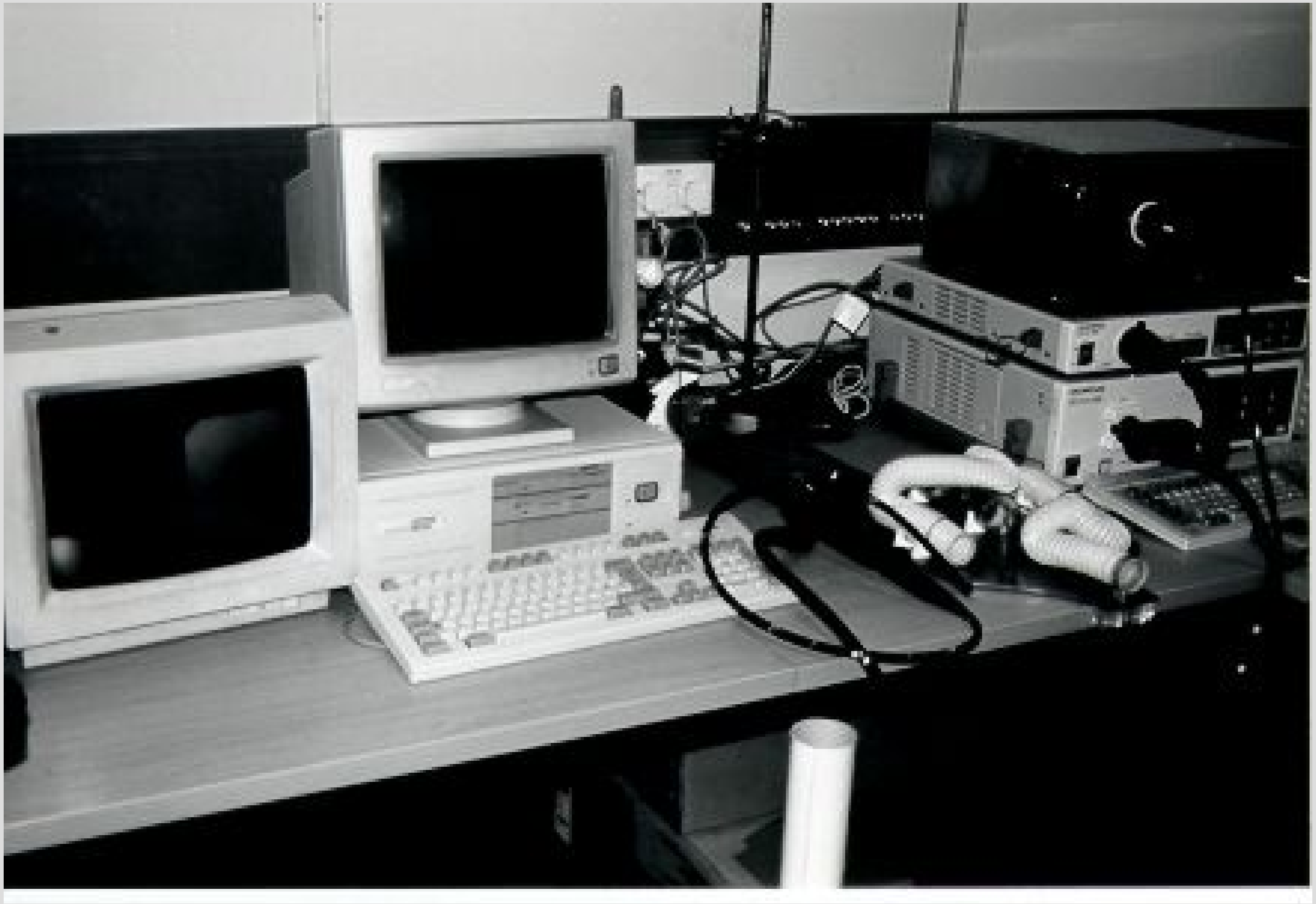


Olivier.Aycard@imag.fr
esucar@inaoep.mx

BN for endoscopy (partial)

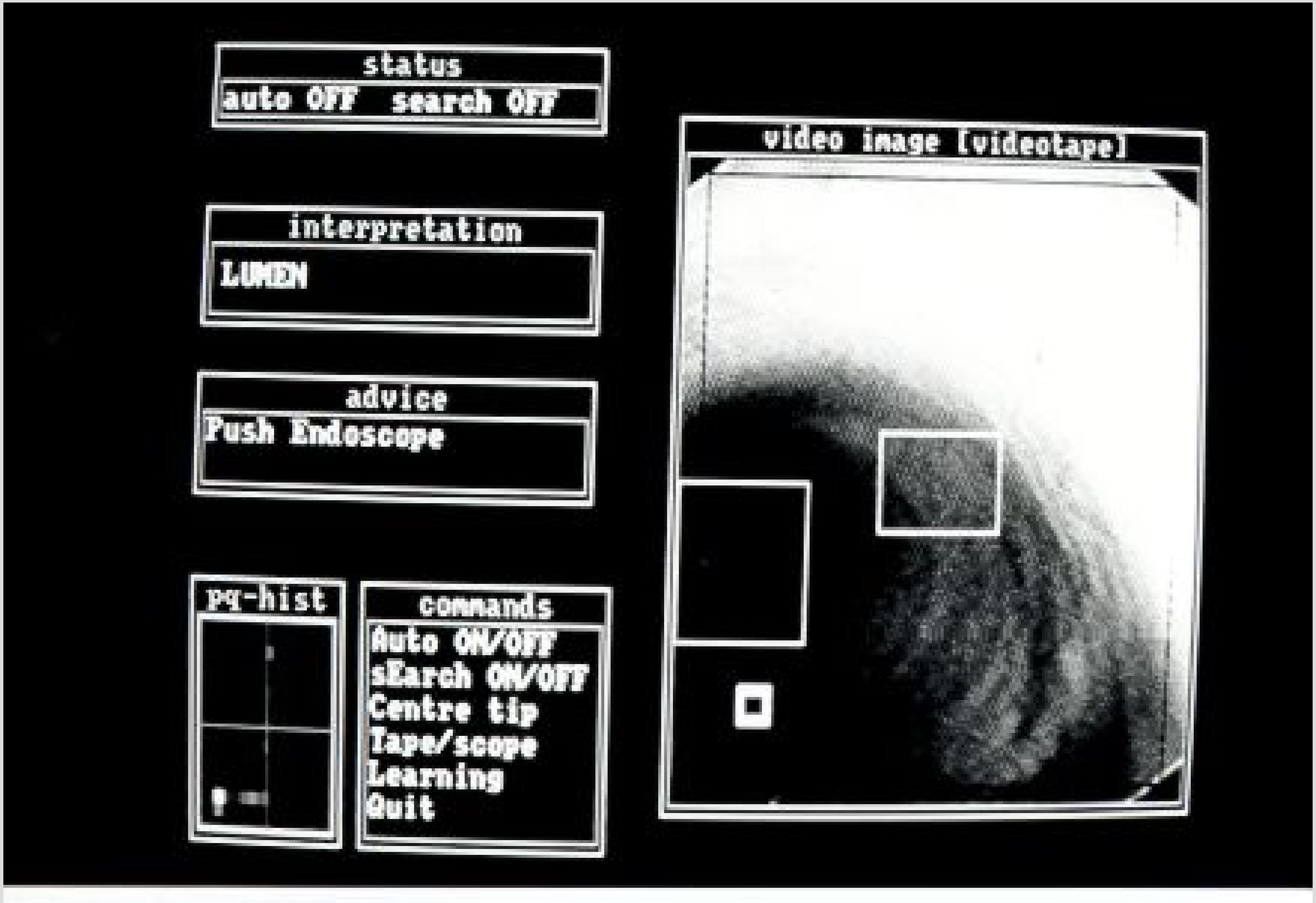


Semi-automatic Endoscope



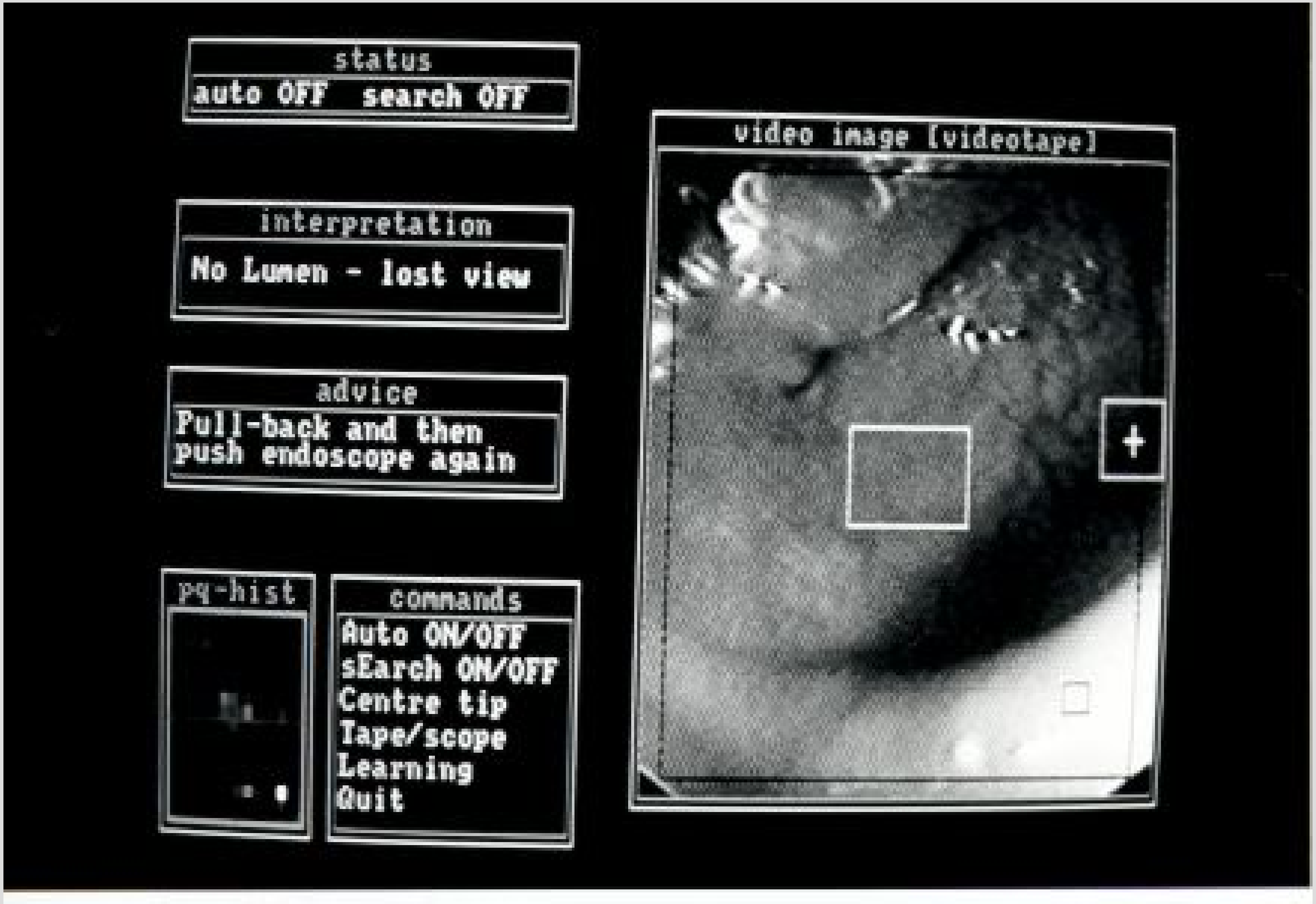
Olivier.Aycard@imag.fr
esucar@inaoep.mx

Endoscopy navegation system



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Endoscopy navegation system

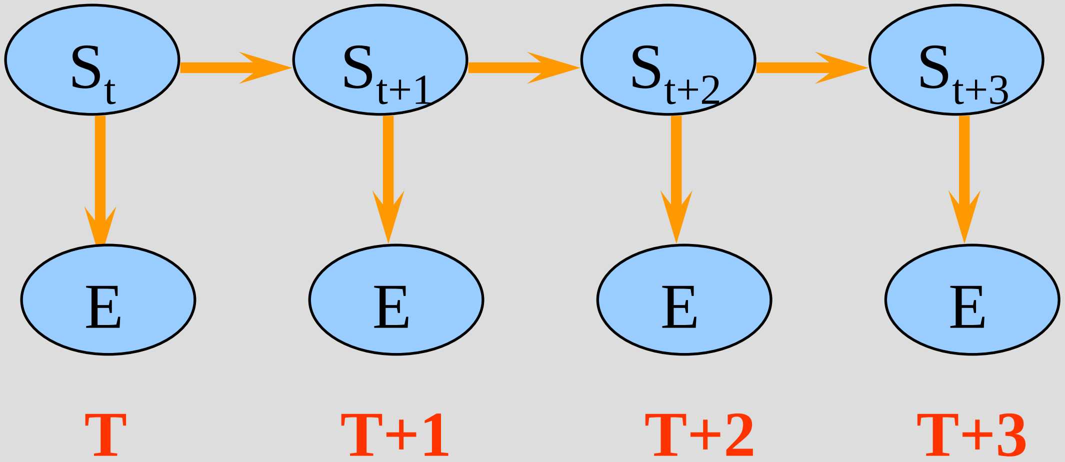


Olivier.Aycard@imag.fr
esucar@inaoep.mx

Dynamic Bayesian networks (DBN)

- BN for modeling temporal processes
- A “static BN” is repeated at each time (discrete time)
- Dependencies (arcs) between temporal slices (Markov assumption)
- Dependencies and parameters between time slices are repeated (Stationary assumption)
- Hidden Markov models (HMMs) are a special case of DBN

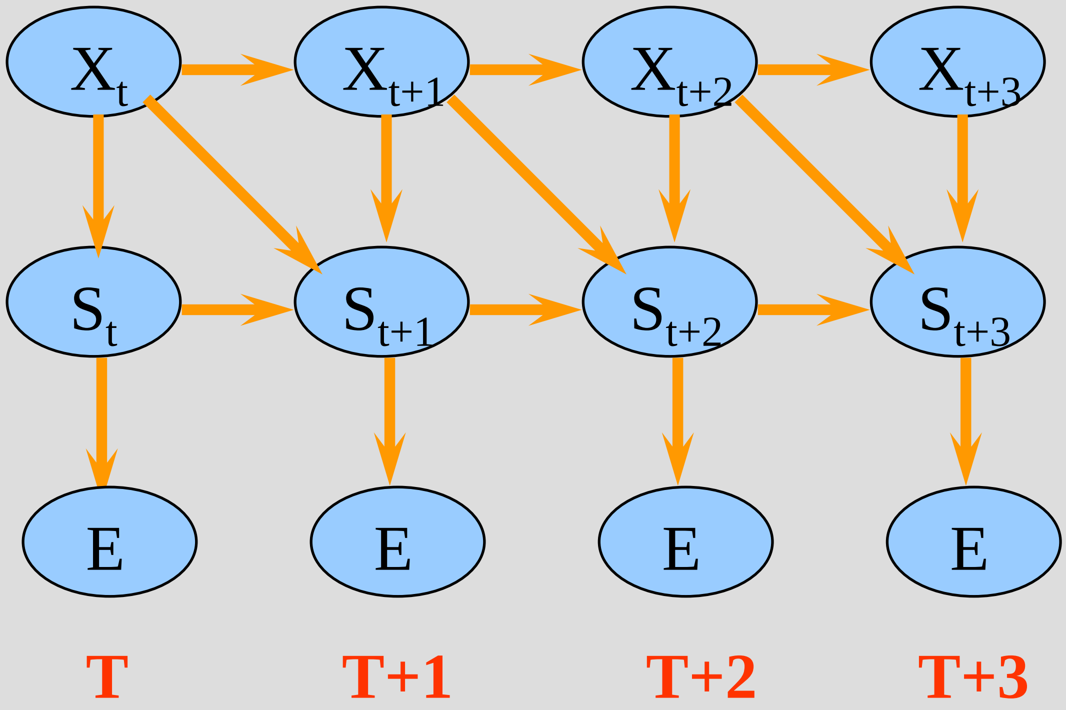
Example of a Dynamic BN (HMM)



Parameters (stationary):

- Initial state probabilities: $P(S_t)$
- State transition probabilities: $P(S_{t+1}|S_t)$
- Observation probabilities: $P(E|S_t)$

Other dynamic BN



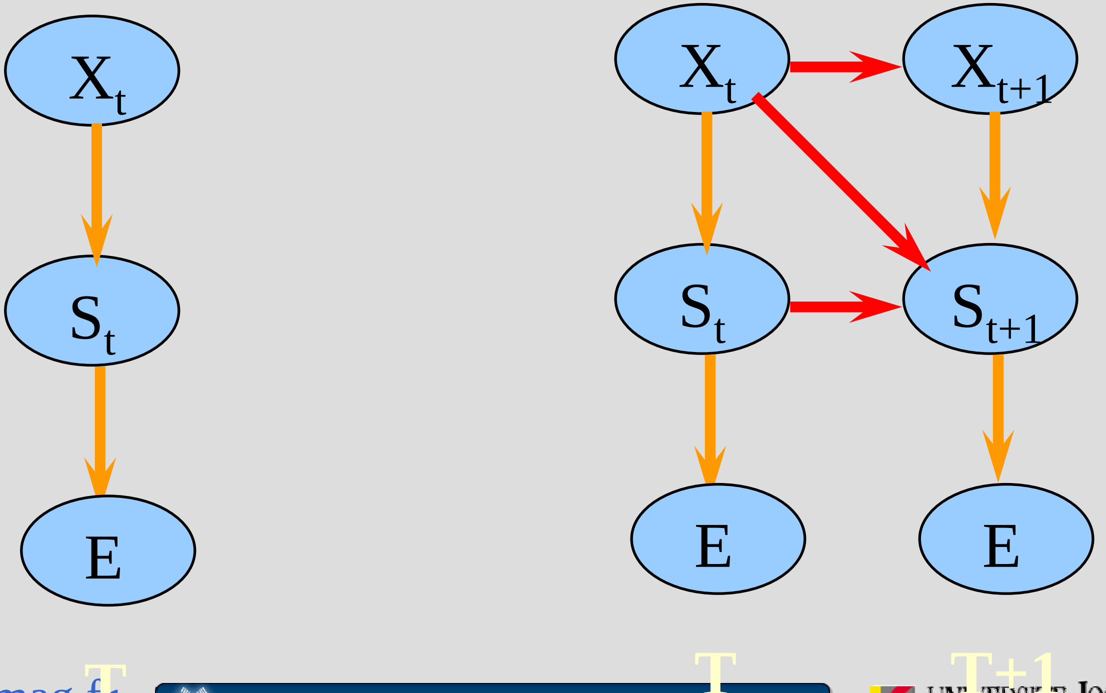
Olivier.Aycard@imag.fr
esucar@inaoep.mx

DBN – Inference

- The same algorithms for BNs apply to DBNs
- But the complexity increases …
- Alternative algorithms:
  - Propagation: as in a static BN (discrete models)
  - Analytic solution: linear models with Gaussian distributions (Kalman filters)
  - Stochastic simulation (P***article filters***): sample the distributions (*particles*) and propagate this to estimate the posterior distributions

DBN – Learning

- Learning is divided into two parts:
  - Learn the "static" structure and parameters
  - Learn the "transition" links and parameters



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Summary of Bayesian Techniques

- The Bayesian approach combines prior knowledge (a priori probability) with evidence (likelihood) based on Bayes theorem
- Graphical models allow for an efficient and clear representation of probability distributions based on dependency & independency relations
- There are different type of PGMs, we have revised 3:
  - Bayesian classifiers
  - Bayesian networks
  - Dynamic Bayesian networks

Summary of Bayesian Techniques

- The Naïve Bayesian classifier makes the simplifying assumption that the attributes are conditionally independent given the class (it can be improved by feature selection and attribute elimination/combination, or by adding dependencies between attributes – TAN, BAN)
- Bayesian networks are DAGs that represent efficiently a joint probability distribution, and are extended for dynamic domains as DBNs
- There are several applications in vision and robotics:
  - Skin detecion, object recognition, segmentation
  - Activity and gesture recognition
  - Sensor fusion, localization
  - SLAM

Content

- Fundamentals of Bayesian Techniques (E. Sucar)
- Bayesian Filters (O. Aycard)
  - Definition & interests
    - Definition
    - Inference
    - Pratical problems
  - Implementations
- Research activities in Vision (E. Sucar)
- Research activities in Perception (O. Aycard)

Bayesian Filters

- Filtering is the problem of sequentially estimating the states of a system as a set of actions and observations become available on-line (one or several sensors)

- S_t : state at time t
- O_t : observation at time t
- A_t : action at time t

- Hypothesis :
  - Order 1 Markov model
    - $P(S_t \mid S_{t-1}\ A_t)$: dynamic model
  - Sensor model
    - $P(O_t \mid S_t)$: sensor model

- Goal : compute *posterior* distribution $P(S_t \mid O_{0:t}\ A_{1:t})$

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Bayesian Filters

- Variables :

  $S_{0:T}$ $O_{0:T}$ $A_{1:T}$

- Decomposition : $P(S_{0:T}, O_{0:T}, A_{1:T}) = P(S_0) \times P(O_0 | S_0) \times$

  $\prod_{t=1}^{T} P(A_t) \times P(S_t | S_{t-1}, A_t) \times P(O_t | S_t)$

- Parametrical forms :

$P(S_t | S_{t-1}, A_t)$: dynamic model

$P(O_t | S_t)$: sensor model

Utilization

$$P(S_t | A_{1:t}, O_{0:t})$$

Recursive Inference Question

$$P\left(S_T \middle| O_{0:T}, A_{1:T}\right) = \frac{1}{Z} P\left(S_T, O_{0:T}, A_{1:T}\right)$$

$$P\left(S_T \middle| O_{0:T}, A_{1:T}\right) = \frac{1}{Z} \int_{S_{T-1}} P\left(S_T, S_{T-1}, O_{0:T}, A_{1:T}\right)$$

$$P\left(S_T \middle| O_{0:T}, A_{1:T}\right) = \frac{1}{Z} \int_{S_{T-1}} P\left(O_T \middle| S_T, S_{T-1}, O_{0:T-1}, A_{1:T}\right) \times P\left(S_T, S_{T-1}, O_{0:T-1}, A_{1:T}\right)$$

$$P\left(S_T \middle| O_{0:T}, A_{1:T}\right) = \frac{1}{Z} P\left(O_T \middle| S_T\right) \int_{S_{T-1}} P\left(S_T \middle| S_{T-1}, O_{0:T-1}, A_{1:T}\right) \times P\left(S_{T-1}, O_{0:T-1}, A_{1:T}\right)$$

$$P\left(S_T \middle| O_{0:T}, A_{1:T}\right) = \frac{1}{Z} P\left(O_T \middle| S_T\right) \int_{S_{T-1}} P\left(S_T \middle| S_{T-1}, A_T\right) \times P\left(S_{T-1} \middle| O_{0:T-1}, A_{1:T}\right) \times P\left(O_{0:T-1}, A_{1:T}\right)$$

$$P\left(S_T \middle| O_{0:T}, A_{1:T}\right) = \frac{1}{Z'} P\left(O_T \middle| S_T\right) \int_{S_{T-1}} P\left(S_T \middle| S_{T-1}, A_T\right) \times P\left(S_{T-1} \middle| O_{0:T-1}, A_{1:T}\right)$$

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Recursive Inference

Initialization :

$$P(S_0|O_0) = \frac{1}{Z} P(S_0) \times P(O_0|S_0)$$

Integral approximation

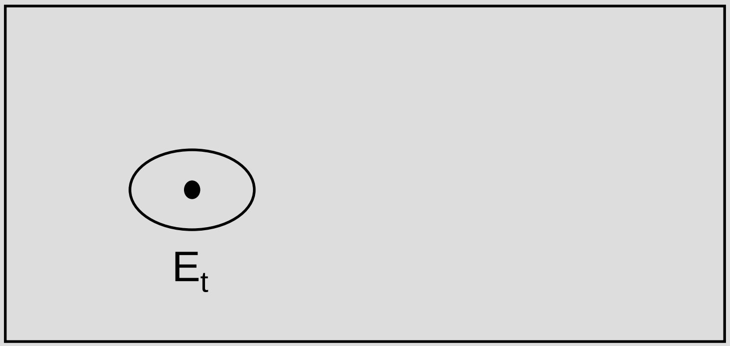
$$P(S_T|O_{0:T-1}, A_{1:T}) = \int_{S_{T-1}} P(S_T|S_{T-1}, A_T) \times P(S_{T-1}|O_{0:T-1}, A_{1:T})$$

Estimation : Confrontation observation - prediction

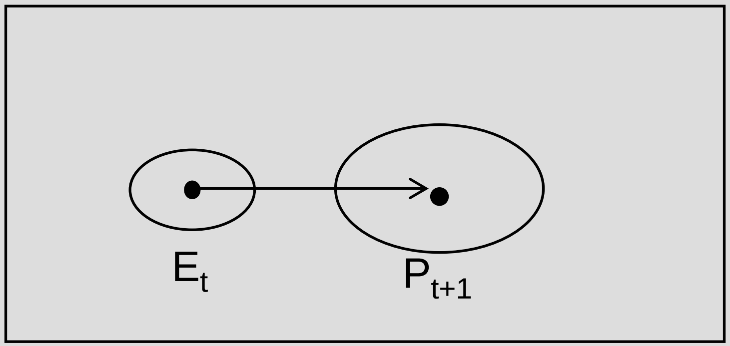
$$P(S_T|O_{0:T}, A_{1:T}) = \frac{1}{Z} P(O_T|S_T) \times P(S_T|O_{0:T-1}, A_{1:T})$$

Olivier.Aycard@imag.fr
esucar@inaoep.mx

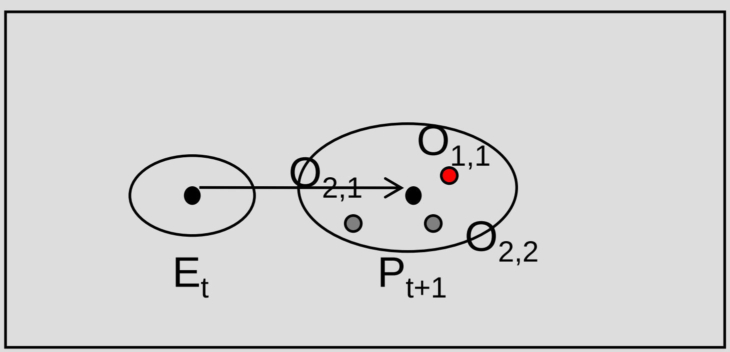
Sequential estimation or Tracking



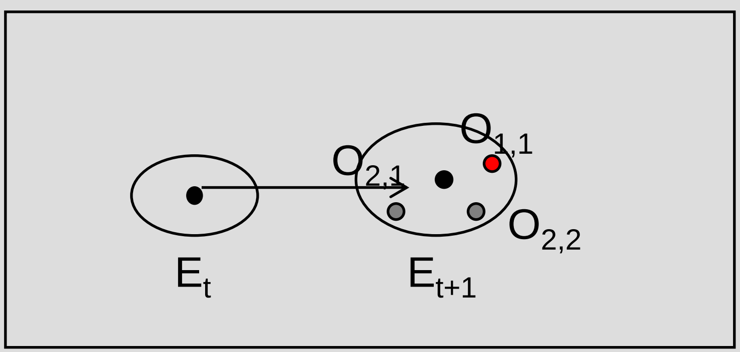
Estimation of position at time t



Prediction of position at time t+1



Observations at time t+1



Estimation of position at time t+1

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Practical problems

- Only a conceptual solution
  - Integrals are seldom intractable
- Practical solutions are only possible under some hypothesis:
  - Discretization of state-space:
    - Markov chain, Hidden Markov models, Markov localization or discrete bayesian filters
  - Representation of state space by a gaussian
    - Kalman filters, Extended Kalman filters
  - Representation of state-space by particles:
    - Particle filters;
- Different type of filters:
  - Markov chain: no actions, no observations;
  - Hidden Markov models: no actions;
  - Markov localization, Kalman filters, particle filters: actions and observations;