Content

- Fundamentals of Bayesian Techniques (E. Sucar)
- Bayesian Filters (O. Aycard)
  - Definition & interests
  - Implementations
    - Hidden Markov models
      - Definition
      - Inference
      - Interests
      - Example
      - Learning phase
      - Recognition phase
      - Conclusion
    - Discrete Bayesian Filters or Markov localization
    - Kalman filters
    - Particle filters
- Research activities in Vision (E. Sucar)
- Research activities in Perception (O. Aycard)

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

Hidden Markov models

- Hidden Markov models is an implementation of Bayesian Filters
- State space is discrete: described by N states
- Specific learning phase: Baum-Welch algorithm
- Specific recognition phase: Viterbi algorithm

- S_t : state at time t
- O_t : observation at time t
- no actions

- Hypothesis :
  - Order 1 Markov model
    - $P(S_t \mid S_{t-1})$: dynamic model
  - Sensor model
    - $P(O_t \mid S_t)$: sensor model
- Goal : compute $P(S_{0:T} \mid O_{0:T})$

Hidden Markov models

- Pattern recognition:
  - Given a sequence of data (ie, observations), recognize the associate pattern;
- Prediction:
  - Given a sequence of data (ie, observations), recognize the associate pattern and predict the end of the sequence;
- Several applications:
  - Speech recognition;
  - Handwriting recognition;
  - Finance;
  - Vision;
  - Robotics;
  - Interpretation of complex signal;
  - Bio-informatic

Hidden Markov models

definition

- S: a finite set of N states: $S = \{1, 2, \ldots, N\}$;
- π: a vector of initial probabilities over S: $\pi_i = P(S_0 = i), 1 \leq i \leq N$;
- A: a matrix of probabilities of transitions over SxS: $a_{ij} = P(S_t = j \mid S_{t-1} = i), 1 \leq i, j \leq N$.
- V: a finite set of M observations: $V=\{1,2,\ldots,M\}$;
- B: a matrix of probabilities of observations over state: $b_i(k) = P(O_t = V_k \mid S_t = i)$

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

Inference in HMM

- Variables

  $O_O, \ldots, O_T, S_O, \ldots, S_T$ **no actions**

- Decomposition

$$P(O_0,\ldots,O_T,S_0,\ldots,S_T) = P(S_0)\times P(O_0 \mid S_0)\times \prod_{t=1}^{T} P(S_t \mid S_{t-1})\times P(O_t \mid S_t)$$

- Parametric forms

$$P(S_t \mid S_{t-1}) \leftarrow \textit{Histograms}$$

$$P(O_t \mid S_t) \leftarrow \text{Gaussians}$$

Learning using the BW algorithm: **a specific learning algorithm**

Utilization : *iterative* question

$P(S_{0:T} \mid O_{0:T})$

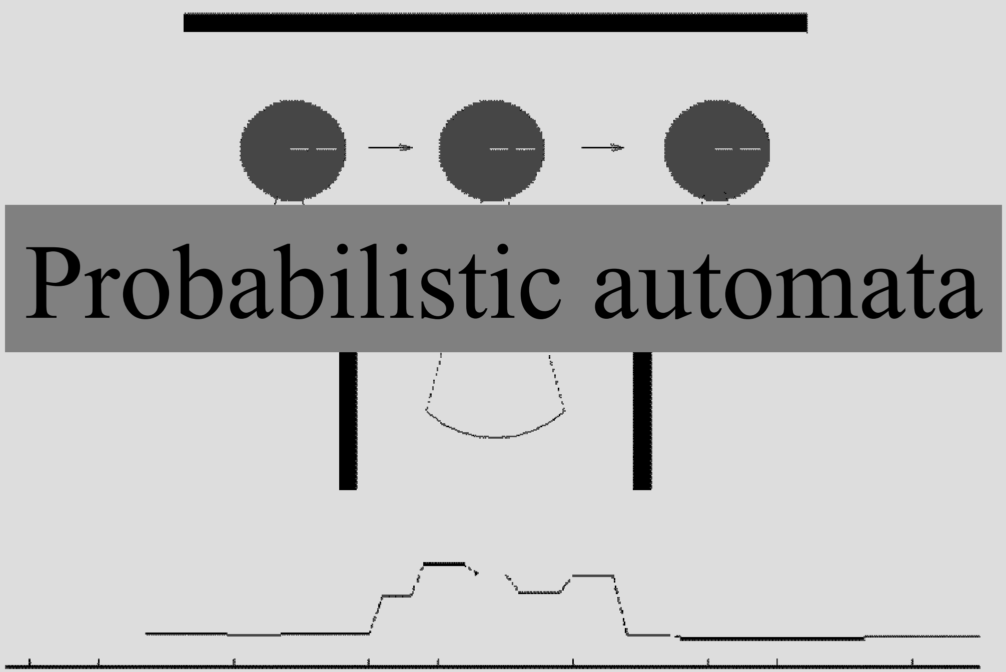
Olivier.Aycard@imag.fr
esucar@inaoep.mx

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA
Luis Enrique Erro # 1, Tonantzintla, Puebla, México Apartado Postal 51 y 216, 72000 Tel: (222) 266-31-00 Fax: (222) 247-22-31

UNIVERSITE JOSEPH FOURIER

Example

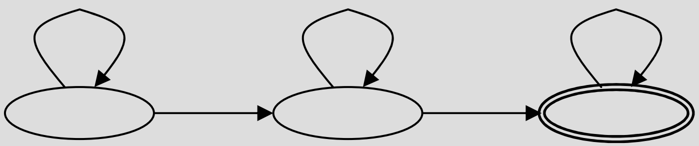
- A mobile robot is moving in a corridor;
- It is equipped with a sensor;
- Given a sequence of noisy sensor data, we want to recognize if the robot is in front of a T-intersection or a wall



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

Hidden Markov models

example: T-intersection model λ_T



$$\pi = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad V = \begin{pmatrix} \text{important increase (+2)} \\ \text{low increase (+1)} \\ \text{quasi null variation (0)} \\ \text{low decrease (-1)} \\ \text{important decrease (-2)} \end{pmatrix}$$

$$A = \begin{pmatrix} 0,28 & 0,72 & 0 \\ 0 & 0,32 & 0,68 \\ 0 & 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 0,17 & 0,06 & 0,06 & 0,71 \\ 0 & 0,42 & 0,32 & 0,26 & 0 \\ 0,72 & 0,11 & 0,11 & 0,06 & 0 \end{pmatrix}$$

Learning phase

goal

- Goal of the learning phase
  - Given a set of sequences of noisy data characterizing a particular pattern (modelized by a HMM), the goal of the learning phase is to estimate the parameters $P(S_t|S_{t-1})$ and $P(O_t|S_t)$ in order to be able to recognize the pattern associated to a new sequence of noisy data;
  - Eventually, the parameters of the model could be given *apriori*.

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

Hidden Markov models

learning of models

Using the maximum likehood criteria:
the most used criteria

$$a_{ij} = \frac{\text{expected number of transitions from state i to state j}}{\text{expected number of transitions from state i}}$$

This is just a count !

$$b_i(k) = \frac{\text{expected number of times in state i observing symbol } v_k}{\text{expected number of times in state i}}$$

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

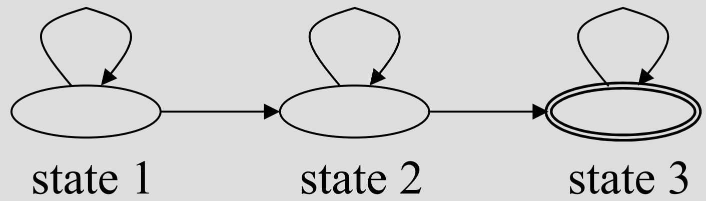
INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

UNIVERSITE JOSEPH FOURIER

Hidden Markov models

example of learning

Model of T-intersection: $\lambda_T = (A_T, B_T, \pi_T, V, N)$



$$A_T = \begin{pmatrix} 0,5 & 0,5 & 0 \\ 0 & 0,5 & 0,5 \\ 0 & 0 & 1 \end{pmatrix} \quad B_T = \begin{pmatrix} 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \\ 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \\ 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \end{pmatrix}$$

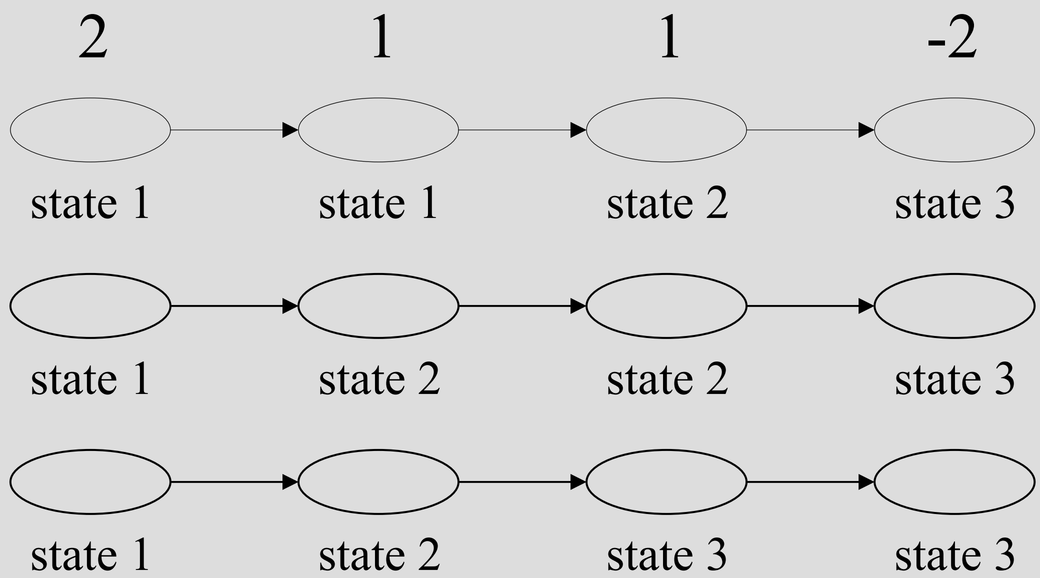
$$\pi_T = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad O^{k_T}_T = \begin{pmatrix} 2 & 0 & -2 & & \\ 2 & 1 & -2 & & \\ 2 & 0 & -1 & -2 & \\ 2 & 1 & 1 & -2 & \\ 2 & -1 & -1 & 0 & -2 \end{pmatrix}$$

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

Hidden Markov models

learning of a sequence of observations

observations



$$A'_T = \begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \qquad B'_T = \begin{pmatrix} 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 4 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \end{pmatrix}$$

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

Hidden Markov models

models after learning

Model of T-intersection: $\lambda_T' = (A_T', B_T', \pi_T, V, N)$

$$A'_T = \begin{pmatrix} 5/18 & 13/18 & 0 \\ 0 & 6/19 & 13/19 \\ 0 & 0 & 5/5 \end{pmatrix} \quad B'_T = \begin{pmatrix} 0 & 3/18 & 1/18 & 1/18 & 13/18 \\ 0 & 8/19 & 6/19 & 5/19 & 0 \\ 13/18 & 2/18 & 2/18 & 1/18 & 0 \end{pmatrix}$$

Model of corridor: $\lambda_C' = (A_C', B_C', \pi_C, V, N)$

$$A'_C = \begin{pmatrix} 4/14 & 10/14 & 0 \\ 0 & 5/15 & 10/15 \\ 0 & 0 & 3/3 \end{pmatrix} \quad B'_C = \begin{pmatrix} 2/14 & 1/14 & 9/14 & 1/14 & 1/14 \\ 3/15 & 3/15 & 6/15 & 3/15 & 0 \\ 3/14 & 5/14 & 4/14 & 2/14 & 0 \end{pmatrix}$$

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

Hidden Markov models

reestimation procedure

- $\prod_k P(O_k \mid \lambda_T') \geq \prod_k P(O_k \mid \lambda_T)$ and $\prod_k P(O_k \mid \lambda_C') \geq \prod_k P(O_k \mid \lambda_C)$
  - The last estimate is more likely than the previous one;

- Reestimation procedure
  - For i from 1 to N do
    - Estimate λ' using O_k (the learning corpus) and λ
    - Each count is weighted by the probability of apparition of the sequence
    - λ <= λ'
  - End_for
- Problem of the choice of N.

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

Hidden Markov models recognition

- Decomposition

$$P(O_0,\ldots,O_T,S_0,\ldots,S_T) = P(S_0)\times P(O_0 \mid S_0)\times\prod_{t=1}^{T} P(S_t \mid S_{t-1})\times P(O_t \mid S_t)$$

Utilization : *iterative* question

$$P(S_0,\ldots,S_T \mid O_0,\ldots,O_T)$$

- Example

  Recognizing if a sequence of 20 observations match to a corridor or to a T-intersection;

  Knowing the sequence of observations: finding the best (highest probability) path over all the possible paths;

  3^{20} possible paths ≈ 3.5 10^9 possible paths

  if 10^6 paths per second => ≈ 58 minutes.

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

Hidden Markov models

recognition of models: Viterbi algorithm

- Given a sequence of observations O, the goal of the recognition is to find the model λ in the set of all the models Λ the probability of recognition (i.e, P(O | λ)) is maximum (given a criteria);
  - Use of a dynamic programming method.

$$\delta_t(i) = \max_{S_1 S_2, \ldots, S_{t-1}} P(S_0, \ldots, S_{t-1}, S_t = i, O_0, \ldots, O_T)$$

$$\delta_0(i) = \pi_i b_i(O_0), 1 \le i \le N$$

$$\delta_{t+1}(j) = \max_{1 \le i \le N} [\delta_t(i) a_{ij}] b_j(o_{t+1}) \quad \begin{matrix} 1 \le t \le T \\ 1 \le j \le N \end{matrix}$$

Similar to the α quantity (max in place of ∑)

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

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

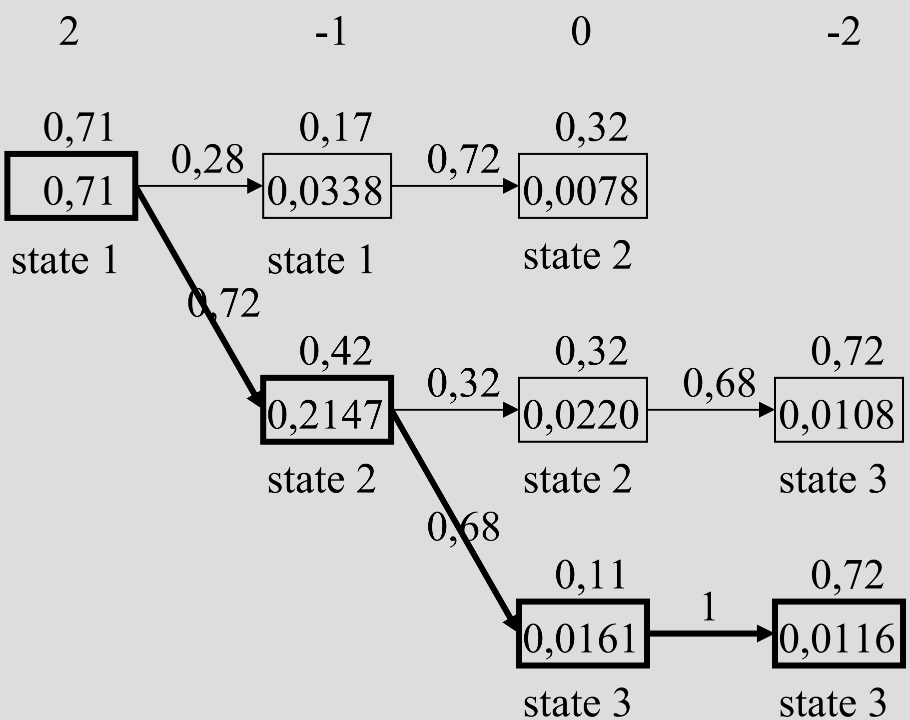
UNIVERSITE JOSEPH FOURIER

Hidden Markov models

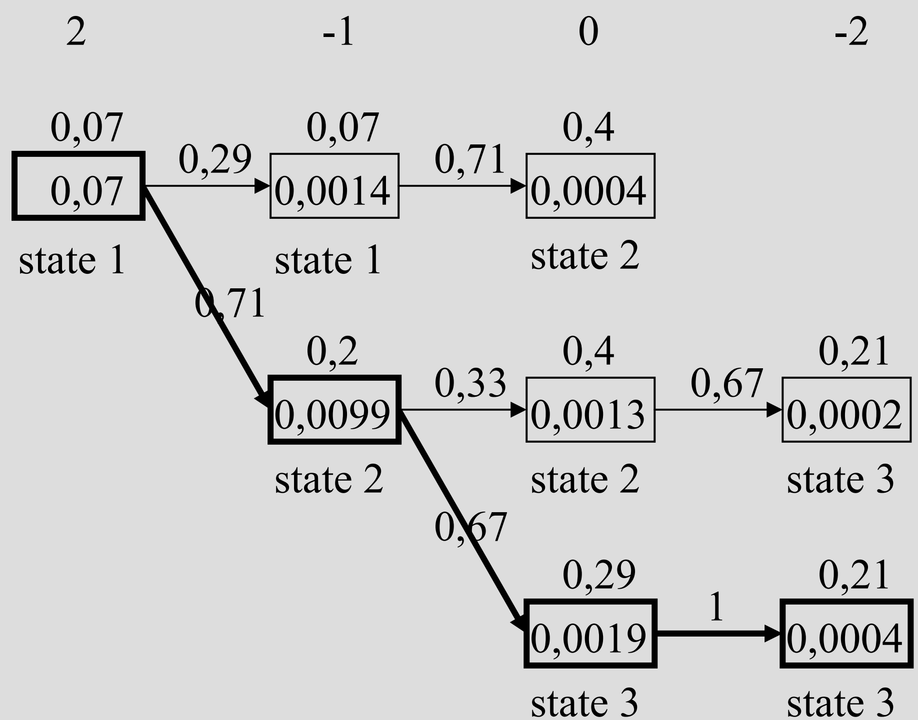
example of recognition

Recognition of the sequence (2 -1 0 -2)

Model of T-intersection



Model of corridor



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

Conclusion

- Powerful tool for pattern recognition
  - Able to recognize pattern using hidden characteristics of the sequence of noisy data
- Particular bayesian filters
  - State space is discrete: described by N states
  - Specific learning phase: Baum-Welch algorithm
  - Specific recognition phase: Viterbi algorithm
- Some areas of application
  - Speech recognition, Handwriting recognition, Finance, Vision, Robotics, Interpretation of complex signal, Bio-informatics;
- Basic model for pattern recognition tools:
  - Factorial Hidden Markov Models, Hierarchical Markov Models, Hidden Markov Decision Trees

Research in the e-Motion group related to HMM

- Past research activities:
  - owerful tool for pattern recognition
  - Able to recognize pattern using hidden characteristics of the sequence of noisy data

- Current research activities:
  - Classification of driver behavior in a car (contract with Toyota Europe)

  - State space is discrete: described by N states
  - Specific learning phase: Baum-Welch algorithm
  - Specific recognition phase: Viterbi algorithm

- Some areas of application
  - Speech recognition, Handwriting recognition, Finance, Vision, Robotics, Interpretation of complex signal, Bio-informatics;

- Basic model for pattern recognition tools:
  - Factorial Hidden Markov Models, Hierarchical Markov Models, Hidden Markov Decision Trees

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

Content

- Fundamentals of Bayesian Techniques (E. Sucar)
- Bayesian Filters (O. Aycard)
  - Definition & interests
  - Implementations
    - Hidden Markov models
    - Discrete Bayesian Filters or Markov localization
      - Definition
      - Interests
      - Inference
      - Example
      - Conclusion
    - Kalman filters
    - Particle filters
- Research activities in Vision (E. Sucar)
- Research activities in Perception (O. Aycard)

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

Discrete Bayesian Filters or Markov Localization

- Markov Localization is an implementation of Bayesian Filters
- State space is discrete: described by N states

- S_t : state at time t
- Z_t *or* O_t : observation at time t
- A_t : actions at time t

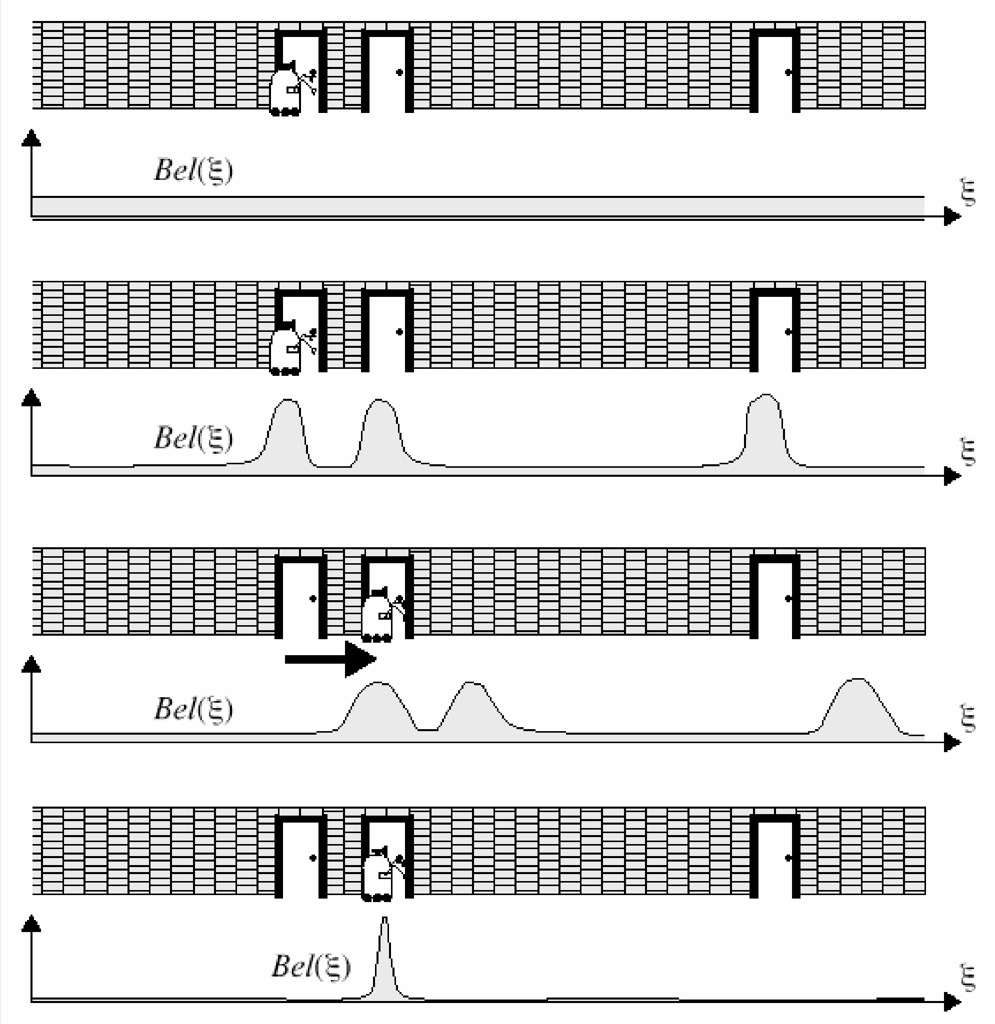
- Hypothesis :
  - Order 1 Markov model
    - $P(S_t \mid S_{t-1}, A_t)$: dynamic model
  - Sensor model
    - $P(Z_t \mid S_t)$: sensor model

- Goal : compute *prior* distribution $P(S_t \mid Z_{0:t}, A_{1:t})$

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

Interest

- Phd of
  Dieter Fox (98)



Inference in Markov localization

- Variables

  $O_0, \ldots, O_T, A_1, \ldots, A_T, S_0, \ldots, S_T$

- Decomposition

$$P(O_0,\ldots,O_T,A_1,\ldots,A_{T-1},S_0,\ldots,S_T) = P(S_0)\times P(O_0 \mid S_0)\times\prod_{t=1}^{T} P(A_t)\times P(S_t \mid S_{t-1},A_t)\times P(O_t \mid S_t)$$

- Parametric forms

$$P(S_t \mid S_{t-1},A_t) \leftarrow \textit{Histograms},\ P(A_t) \leftarrow \textit{Uniform},$$
$$P(O_t \mid S_t) \leftarrow \text{Gaussians or histograms}$$

Apriori or learning

Utilization : *iterative* question

$$P(S_T \mid O_0,\ldots,O_T,A_1,\ldots,A_T)$$

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

Markov localization: assumptions

- State space is discrete
- *Exact* inference is used to compute:

$$P(S_T \mid O_0, \ldots, O_T, A_1, \ldots, A_T)$$

Markov localization: inference

Initialisation:

$$\mathrm{P}(S_0 \mid O_0) = \frac{1}{Z}\mathrm{P}(S_0, O_0)$$

Recursion:

$$P(S_T | O_0,\ldots,O_T, A_1,\ldots,A_T) = \frac{1}{Z'}\sum_{S_{T-1}}\begin{pmatrix}\mathrm{P}(S_{T-1} \mid O_0,\ldots,O_{T-1}, A_1,\ldots,A_{T-1})\times \\ \mathrm{P}(A_T)\times\mathrm{P}(S_T \mid S_{T-1}, A_T)\times\mathrm{P}(O_T \mid S_T)\end{pmatrix}$$

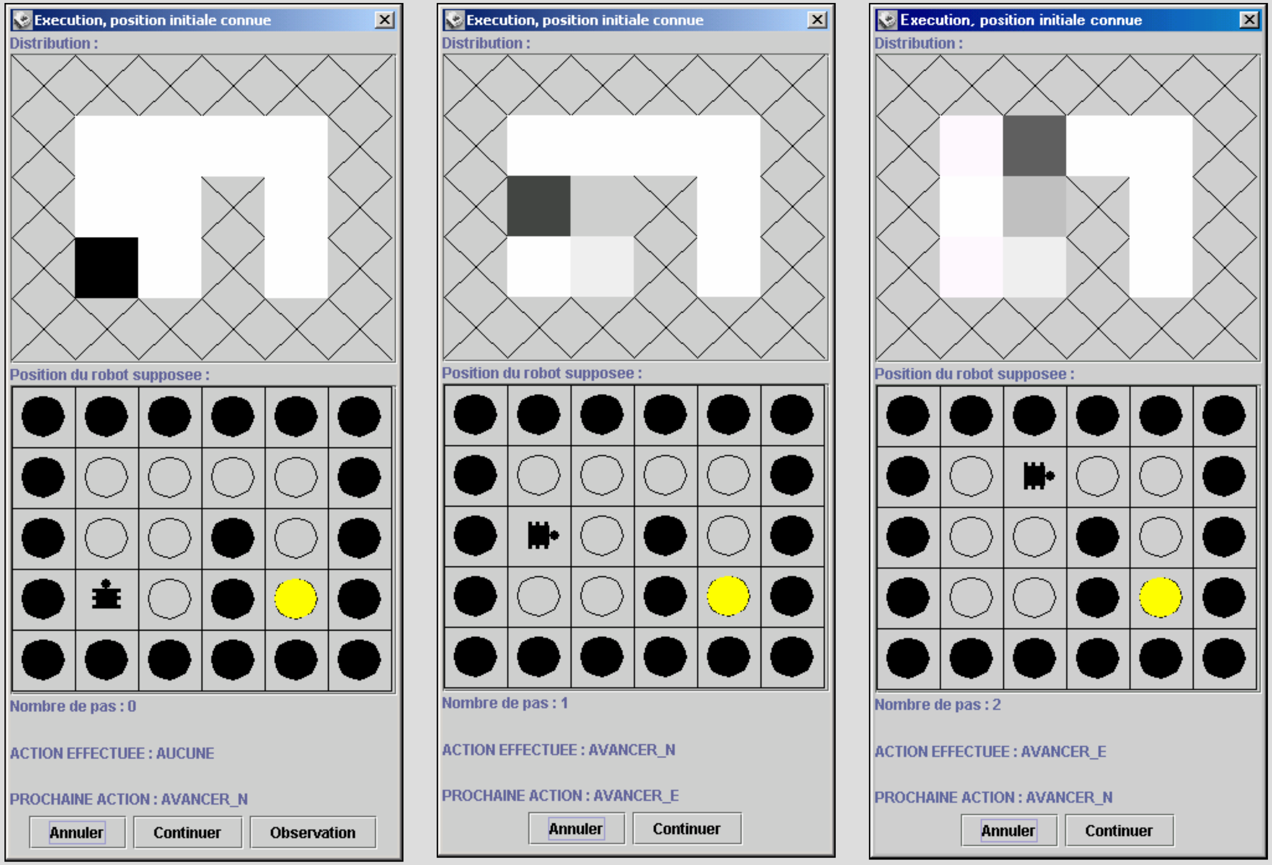
$$P(S_T | O_0,\ldots,O_T, A_1,\ldots,A_T) = \frac{1}{Z'}\mathrm{P}(A_T)\times\mathrm{P}(O_T \mid S_T)\sum_{S_{T-1}}\left(\mathrm{P}(S_{T-1} \mid O_0,\ldots,O_{T-1}, A_1,\ldots,A_{T-1})\times\mathrm{P}(S_T \mid S_{T-1}, A_T)\right)$$

Markov localization: examples

Main hypothesis: the environment is static.

- 3 possible cases for $P(S_0)$:
  - initial position known (Tracking problem);
  - initial position unknown (Global localization);
  - Initial position false (Knidnapping problem).

Initial position known: ultrasonic sensors



a *b* *c*

Initial position unknown: ultrasonic sensors

Conclusion

- Discrete bayesian filters or Markov localization is based on a discretization of the space-states;
- Compute the exact value of $P(S_t|O_{o:t},A_{1:t})$;
- Can handle any $P(S_0)$ (uniform, dirac, wrong value);
- Computation increases exponentially in function of the space-states;
  - In practice, not usable for space-states of dimension >4;

Content

- Fundamentals of Bayesian Techniques (E. Sucar)
- Bayesian Filters (O. Aycard)
  - Definition & interests
  - Implementations
    - Hidden Markov models
    - Discrete Bayesian Filters or Markov localization
    - Kalman filters
    - Particle filters
- Research activities in Vision (E. Sucar)
- Research activities in Perception (O. Aycard)

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

Kalman Filter

- Kalman Filter is an implementation of Bayesian Filters
- State space is represented by a gaussian distribution

- S_t: state at time t
- O_t: observation at time t
- A_t : actions 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 *prior* distribution $P(S_T \mid O_{0:T}, A_{1:T})$

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

Kalman filter : assumptions(1/2)

- S_t is a gaussian distribution represented by $N(\mu_t, \Sigma_t)$
- Dynamic model is linear and associated noise is gaussian
  - $S_t = P_t S_{t-1} + B_t A_t + \varepsilon_t$
    - S_t is a n vector
    - P_t is n x n matrix
    - A_t is a m vector
    - B_t is n x m matrix
    - ε_t is a gaussian distribution represented by $N(0, R_T)$ modelizing the noise associated with the dynamic model

$$P(S_T \mid S_{T-1}, A_T) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det(R_T)}} e^{-\frac{1}{2}\left((S_T - P_T S_{T-1} - B_T A_T)^T R_T^{-1} (S_T - P_T S_{T-1} - B_T A_T)\right)}$$

Kalman filter : assumptions(2/2)

- Sensor model is linear and associated noise is gaussian
  - $O_t = C_t S_t + \delta_t$
  - O_t is a k vector
  - C_t is a k x n matrix
  - δ_t is a gaussian distribution represented by $N(0,Q_t)$ modelizing the noise associated with the observation model

$$P(O_T \mid S_T) = \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det(Q_T)}} e^{-\frac{1}{2}\left((O_T - C_T S_T)^T Q_T^{-1} (O_T - C_T S_T)\right)}$$

- **S_{t+1} is a gaussian distribution represented by $N(\mu_{t+1}, \Sigma_{t+1})$**
  - **Only mean and covariance needed;**
  - **Exact inference using simple linear equations**

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

Kalman filter : equations

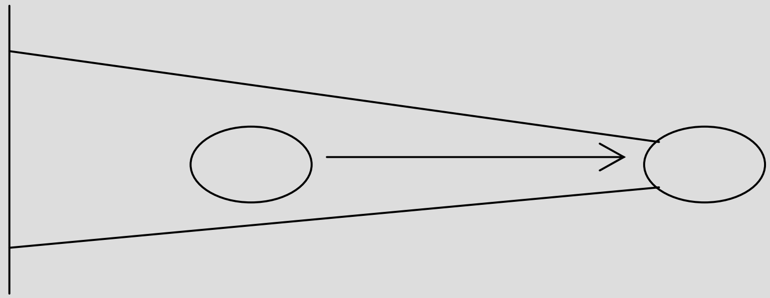
$$\left.\begin{array}{l}\mu_{t+1} = P_t \mu_t + B_t A_t \\ \Sigma_{t+1} = P_t \Sigma_t P_t^T + R_t\end{array}\right\}$$ Prediction

$$K_t = \Sigma_t C_t^T (C^t \Sigma_t C_t^T + Q_t)^{-1}$$ Kalman gain

$$\left.\begin{array}{l}\mu_{t+1} = \mu_{t+1} + K_t(O_t - C_t\, \mu_{t+1}) \\ \Sigma_{t+1} = (I - K_t C_t) \Sigma_{t+1}\end{array}\right\}$$ Estimation

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

Example (1/6)

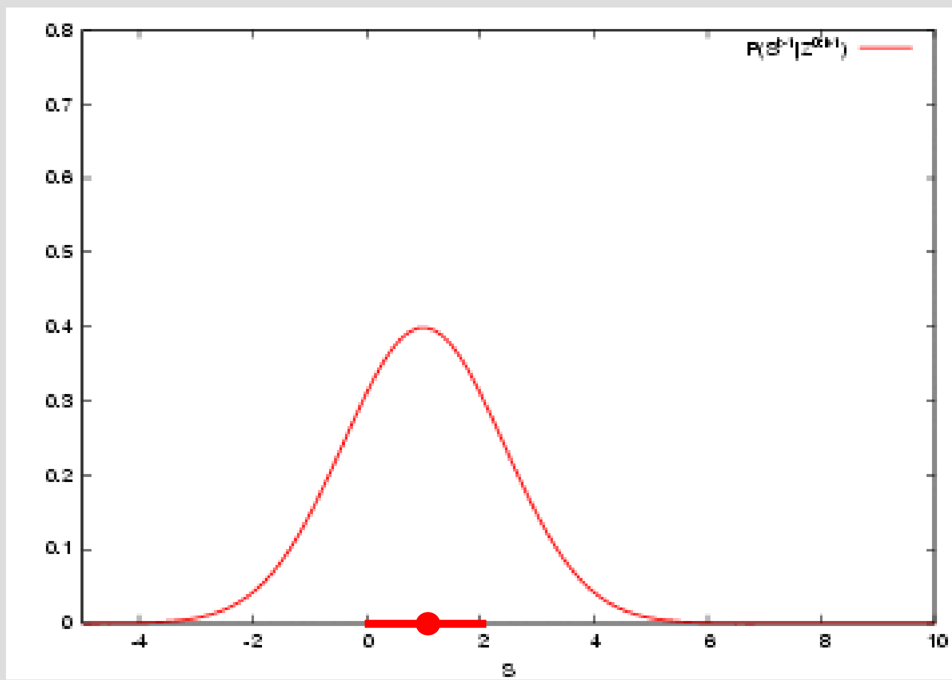


- A mobile robot is initially located at 1 meter from a wall;
- It is moving away form this wall with discrete translation of 3 meters;
- It is equipped with a sensor to measure its distance to the wall.

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

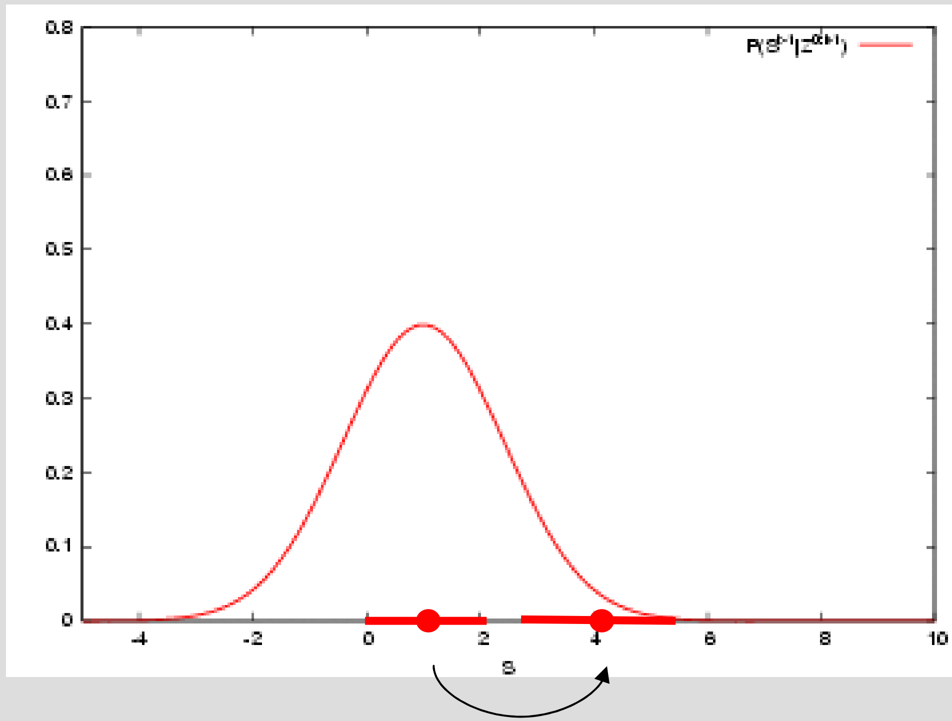
Example: initial stage(2/6)

$\mu_0 = 1$
$\Sigma_0 = 2$



Example (3/6)

$\mu_0 = P_1\mu_0 + B_1A_1 = 1x1 + 1x3 = 4$

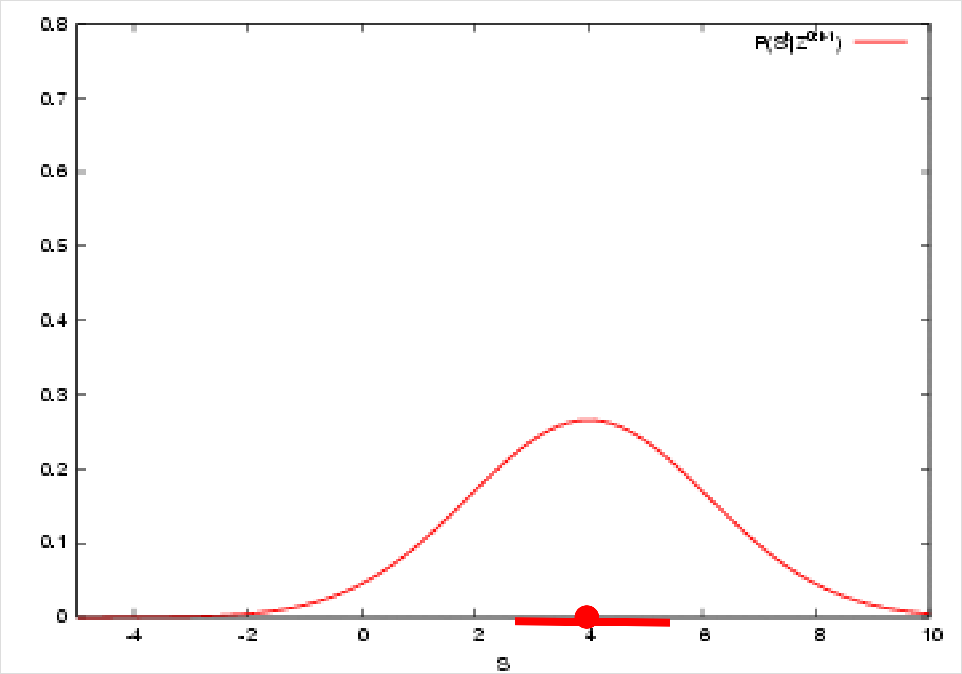


Dynamic model

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

Example(4/6): prediction stage

$$\Sigma_0 = P_1 \Sigma_0 P_1^T + R_1 = 1x2x1 + 2 = 4$$



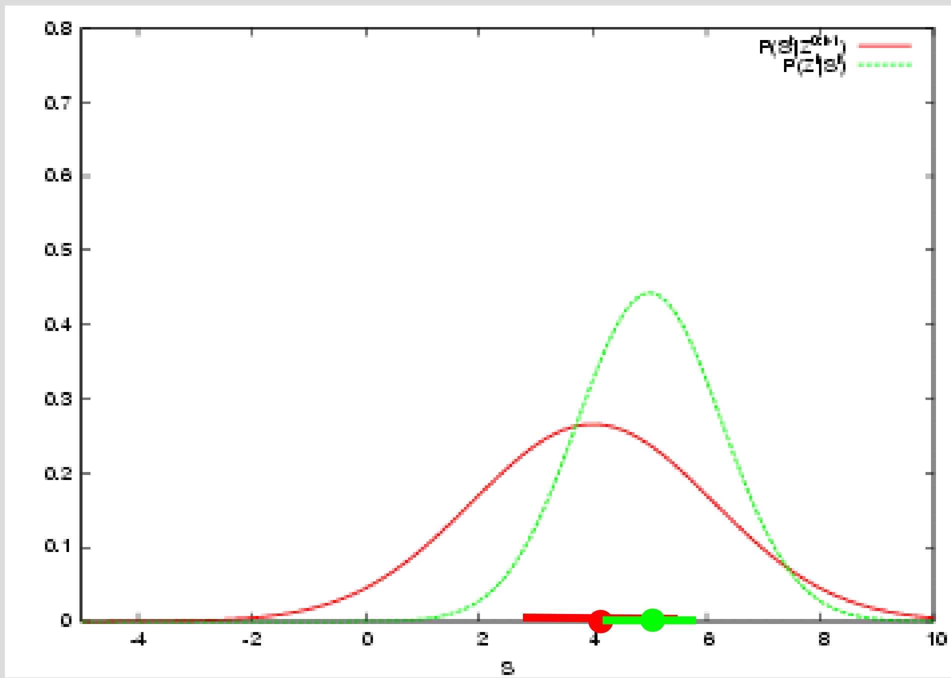
$P(S_1|A_1)$

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

Example (5/6): estimation stage

$O_1 = 5$

$$K_1 = \Sigma_1 C_1^T (C_1 \Sigma_1 C_1^T + Q_1)^{-1} = 4x1(1x4x1+2)^{-1} = 2/3$$



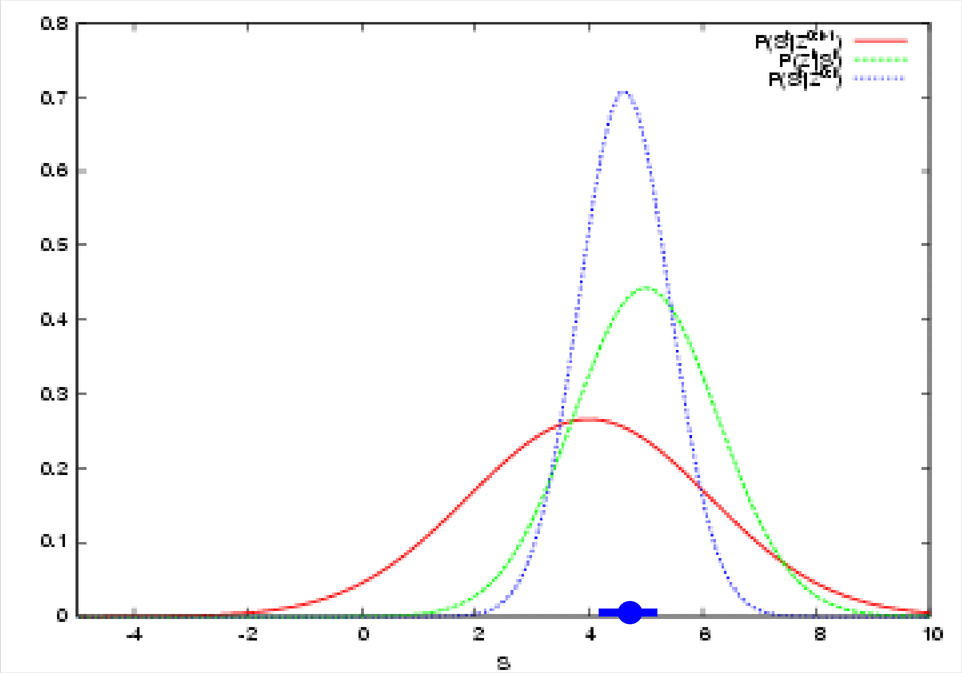
Sensor observation at time t

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

Example (6/6): estimation stage

$\mu_1 = \mu_1 + K_1(O_1 - C_1\mu_1) = 4 + 2/3(5-1x4) = 4+2/3 = 14/3$

$\Sigma_1 = (I-K_1C_1)\Sigma_1 = (1-2/3x1)x4 = 4/3$



Posterior pdf at time t

$P(S_1|A_1,O_1)$

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

Conclusion

- KF limited to *linear* models (sensor and dynamic)
- EKF linearizes the estimation around the current estimate
  - Local linearization using Taylor expansion
  - Similar to KF (ie, gaussian distribution for the state space)
- KF and EKF popular and successful
- KF limited to gaussian noise
- KF can only represent monomodale distributions: only tracking
- Derived models:
  - Unscented Kalman Filter (dynamic model non linear);
  - Multi Hypothesis Kalman Filter.