A Markovian Approach to the Induction of Regular String Distributions

Jérôme Callut, Pierre Dupont

ICGI 2004
October 11th 2004

UCL, Belgium,
Computing science & engineering dept.

HMM Learning Problems

Given a sample resulting from an unknown target model, estimate a model that best explains the observed sample.

The structure is known

$\longrightarrow$ For instance, parameters can be estimated with ***EM***

The structure is unknown

$\longrightarrow$ Structure + parameters have to be estimated

HMM Learning Problems

Given a sample resulting from an unknown target model,
estimate a model that best explains the observed sample.

The structure is known

$\longrightarrow$ For instance, parameters can be estimated with ***EM***

The structure is unknown

$\longrightarrow$ **Structure + parameters have to be estimated**

HMM Learning Problems

Given a sample resulting from an unknown target model, estimate a model that best explains the observed sample.

The structure is known

→ For instance, parameters can be estimated with ***EM***

The structure is unknown

→ **Structure + parameters have to be estimated**

- ***P1*** : Identify at the limit the target model
- ***P2*** : Bayesian framework: trade-off likelihood–model prior
- ***P3*** : Induce a model fitting the dynamics of the target model

HMM Learning Problems

Given a sample resulting from an unknown target model,
estimate a model that best explains the observed sample.

The structure is known

$\longrightarrow$ For instance, parameters can be estimated with ***EM***

The structure is unknown

$\longrightarrow$ **Structure + parameters have to be estimated**

- ***P1*** : Identify at the limit the target model
- ***P2*** : Bayesian framework: trade-off likelihood–model prior
- ***P3*** : **Induce a model fitting the dynamics of the target model**

Approaches to structure induction

1. EM pruning

- Define a surgenerative model (e.g. fully connected graph)
- Estimate parameters with EM
- Prune transitions having small probabilities

$\longrightarrow$ **Poor results in practice.**

2. State merging

- **ALERGIA/RLIPS (1994)**
  - Merge states if their suffix probabilities are close enough.
  - Identification of *structurally deterministic PA*
  - PNFA are identifiable [Denis, Esposito 2004]

  $\longrightarrow$ **ALERGIA: Concerns a restricted class of HMMs**
  $\longrightarrow$ **PNFA identification: Not really tractable algorithm**

- **Stolcke (1994), Schliep (2001)**:
  - Infer an overfitting model
  - Merging/Model Selection is driven by a MAP criterion

  $\longrightarrow$ **High complexity: only applicable on toy problems**

Approaches to structure induction

3. State splitting

- **Orstendorf (1997)**:
  - Initial hand-crafted model is estimated with EM
  - State splitting is driven by the likelihood gain

  $\longrightarrow$ **Restricted to left-to-right HMMs**

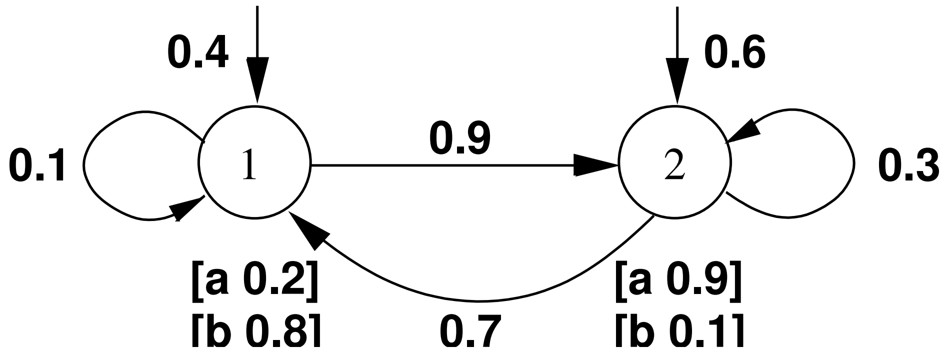
- **Callut, Dupont (2004)**:
  - Initial Model: MC estimated by maximum likelihood
  - State splitting is driven by the sample dynamics

  $\longrightarrow$ **No restriction on model topology**

Agenda

1. HMM: the selected model
2. Partially Observable Markov Models
3. Fundamental quantities of Markov chains
4. Relation between POMM and Markov chains
5. Induction algorithm

HMM: the selected model



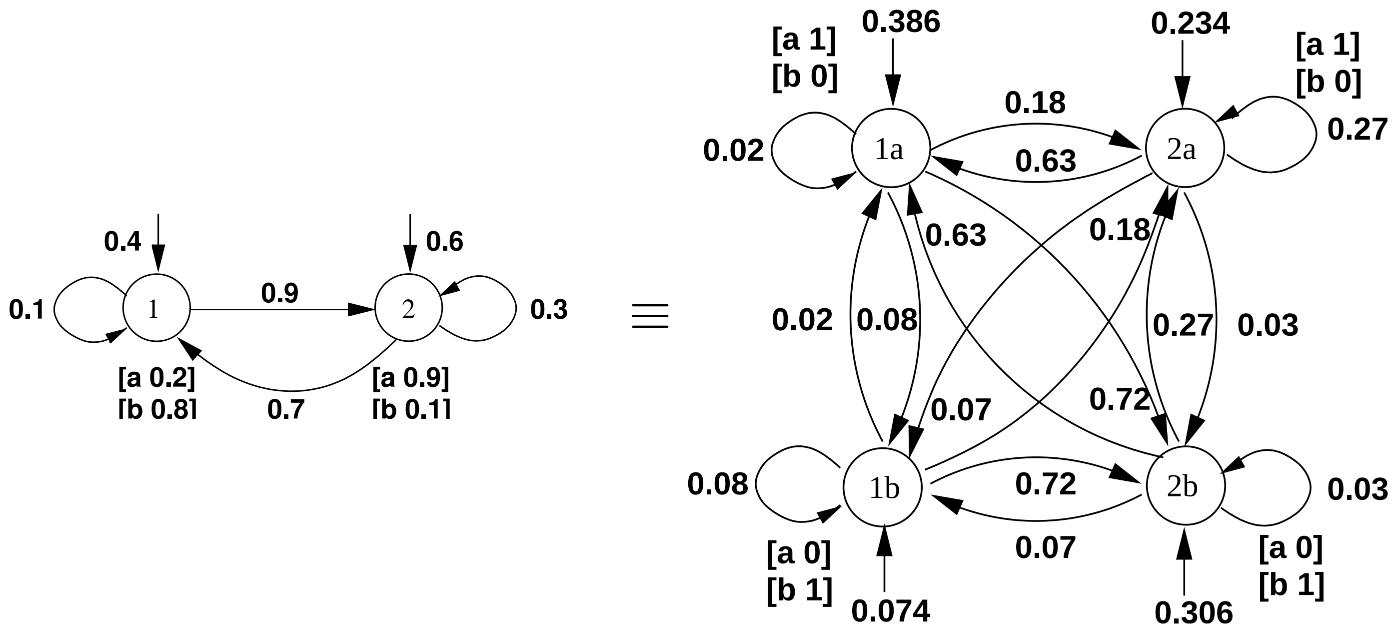
- ***5-tuple:*** $\langle \Sigma, Q, A, B, \iota \rangle$
- ***Structure:*** underlying graph
- ***Parameters:*** transition + emission probabilities
- ***Random walks:*** produce strings defined on an alphabet Σ

Note : No final probabilities are considered.

Partially Observable Markov Models (POMM)

A ***POMM*** is a HMM such that any state emits a ***single letter*** with probability 1.

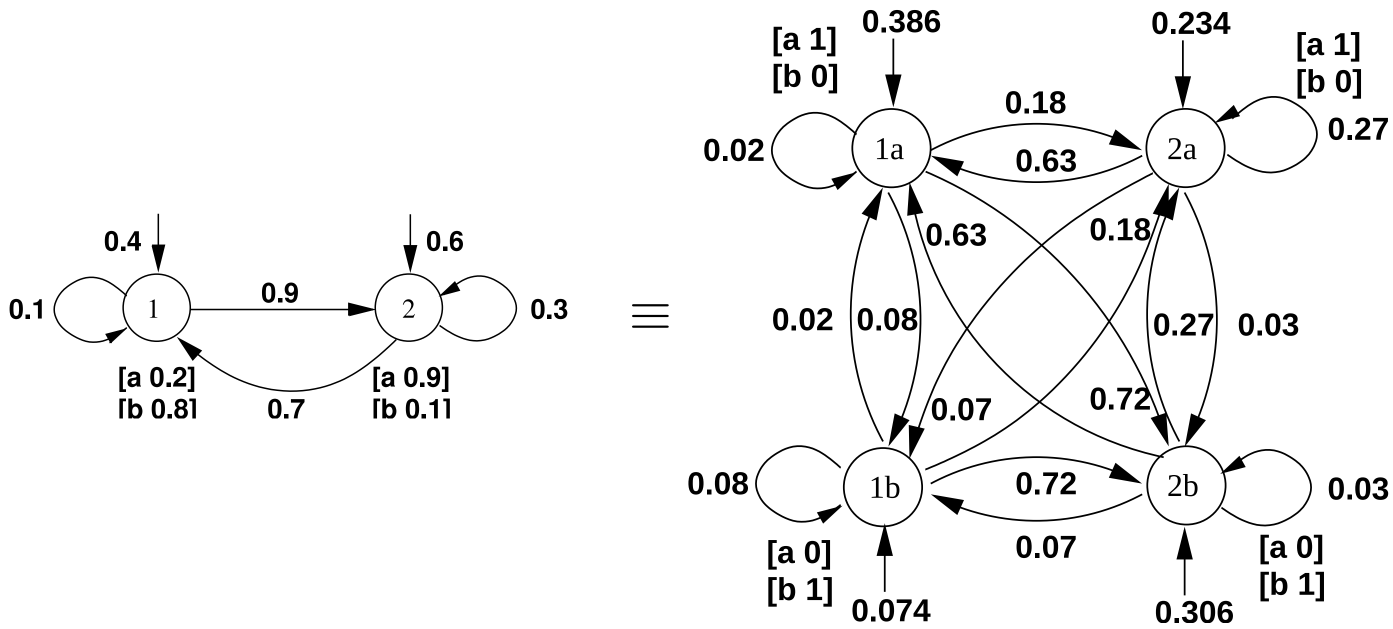
For any HMM there is an ***equivalent*** POMM.



Partially Observable Markov Models (POMM)

A ***POMM*** is a HMM such that any state emits a ***single letter*** with probability 1.

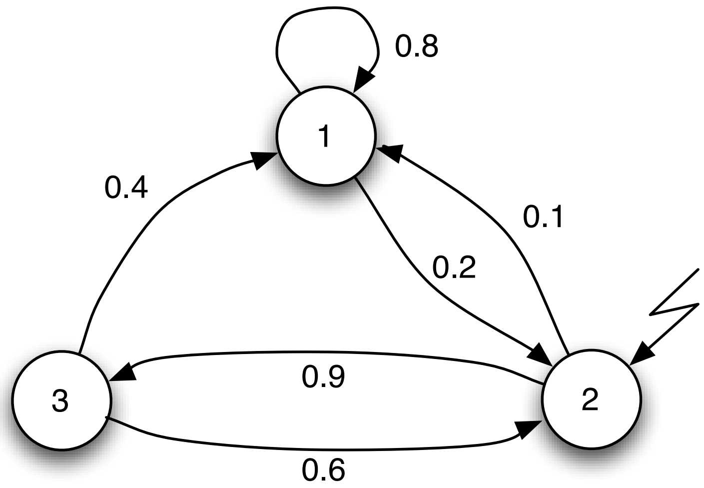
For any HMM there is an ***equivalent*** POMM.



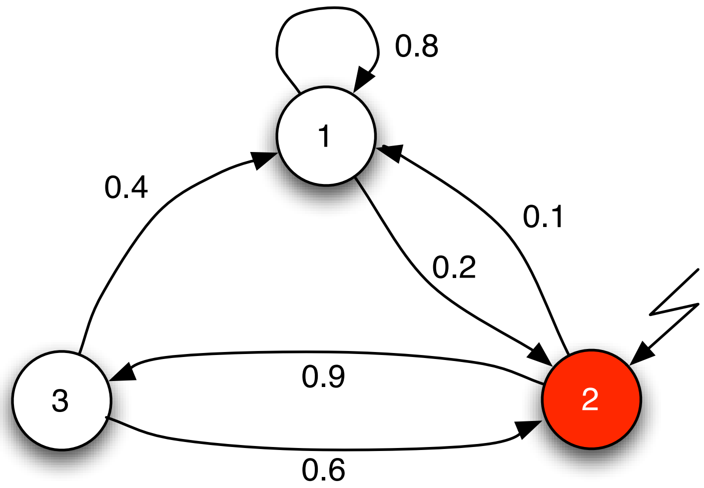
Note:

- Several states can emit the same letter.
- When all states emit distinct letters $\rightarrow$ ***POMM*** $\equiv$ ***Markov chain.***

Fundamental quantities of Markov chains

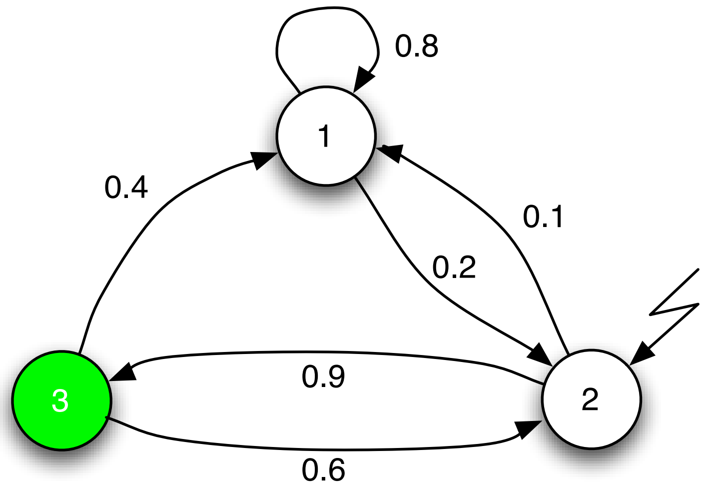


Fundamental quantities of Markov chains



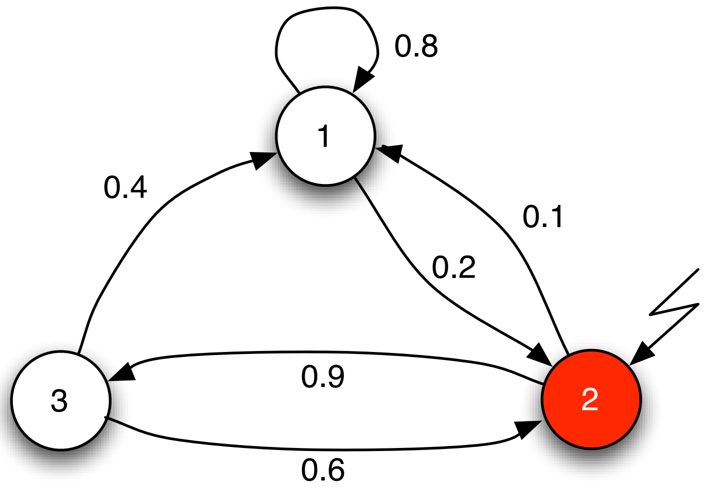
s = 2

Fundamental quantities of Markov chains



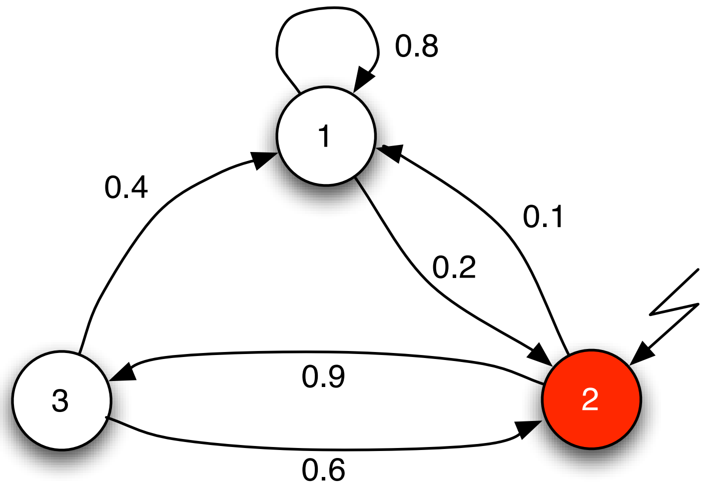
s = 2 3

Fundamental quantities of Markov chains



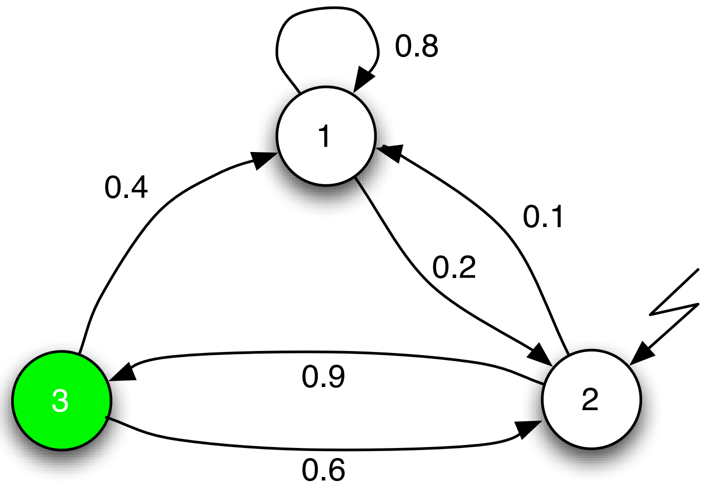
s = 2 3 2

Fundamental quantities of Markov chains



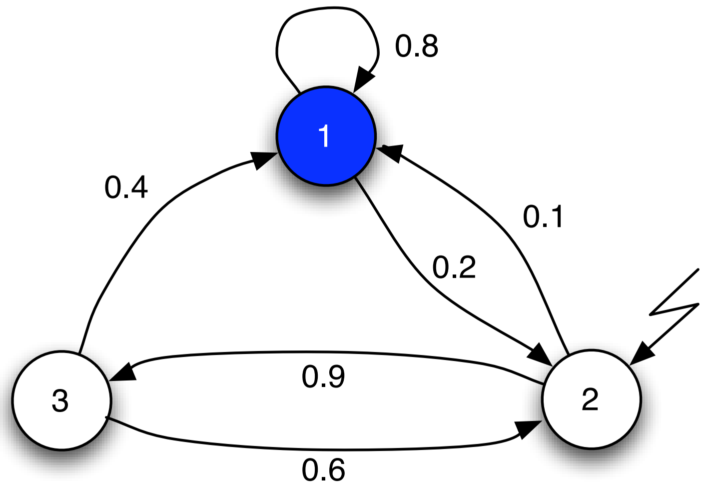
s = 2 3 2 2

Fundamental quantities of Markov chains



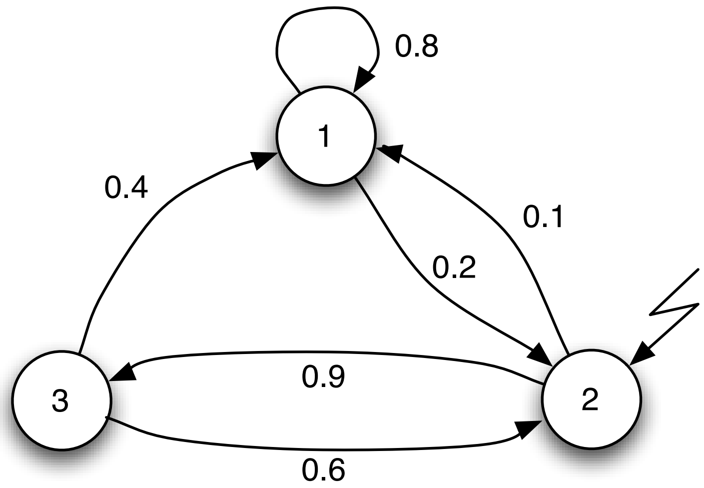
s = 2 3 2 2 3

Fundamental quantities of Markov chains



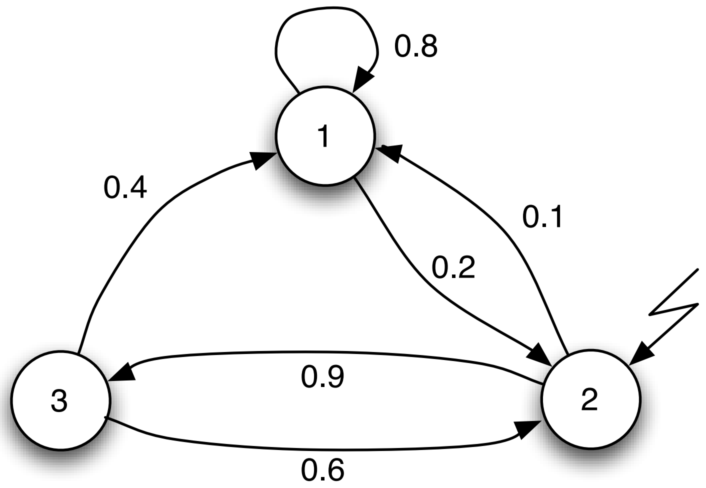
s = 2 3 2 2 3 1

Fundamental quantities of Markov chains



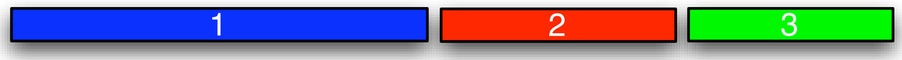
s = 2 3 2 2 3 1 1 1 2 3 2 3 2 1 1 1 1 1 1 2 3 1 1 1 1 1 ...

Fundamental quantities of Markov chains

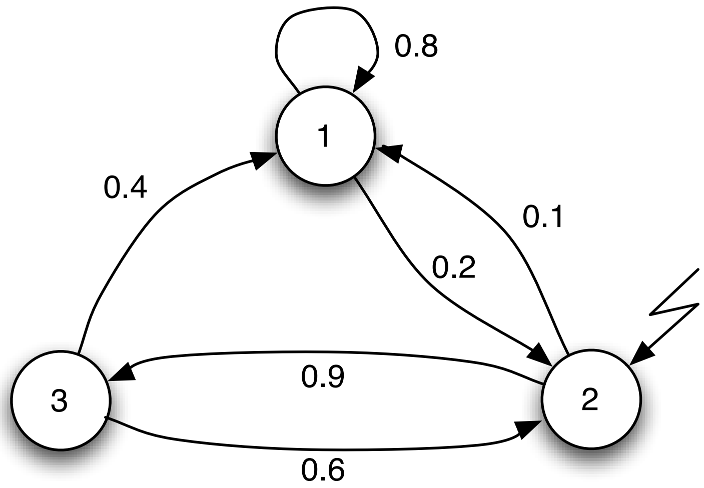


s = 2 3 2 2 3 1 1 1 2 3 2 3 2 1 1 1 1 1 1 2 3 1 1 1 1 1 ...

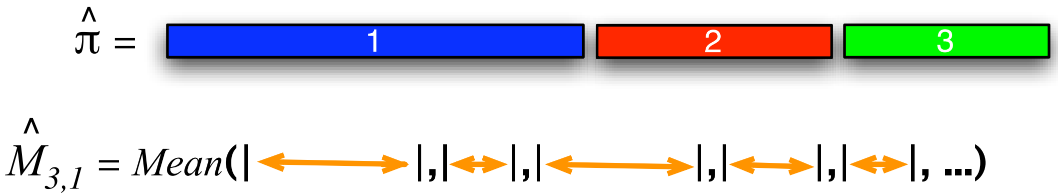
$\hat{\pi} =$



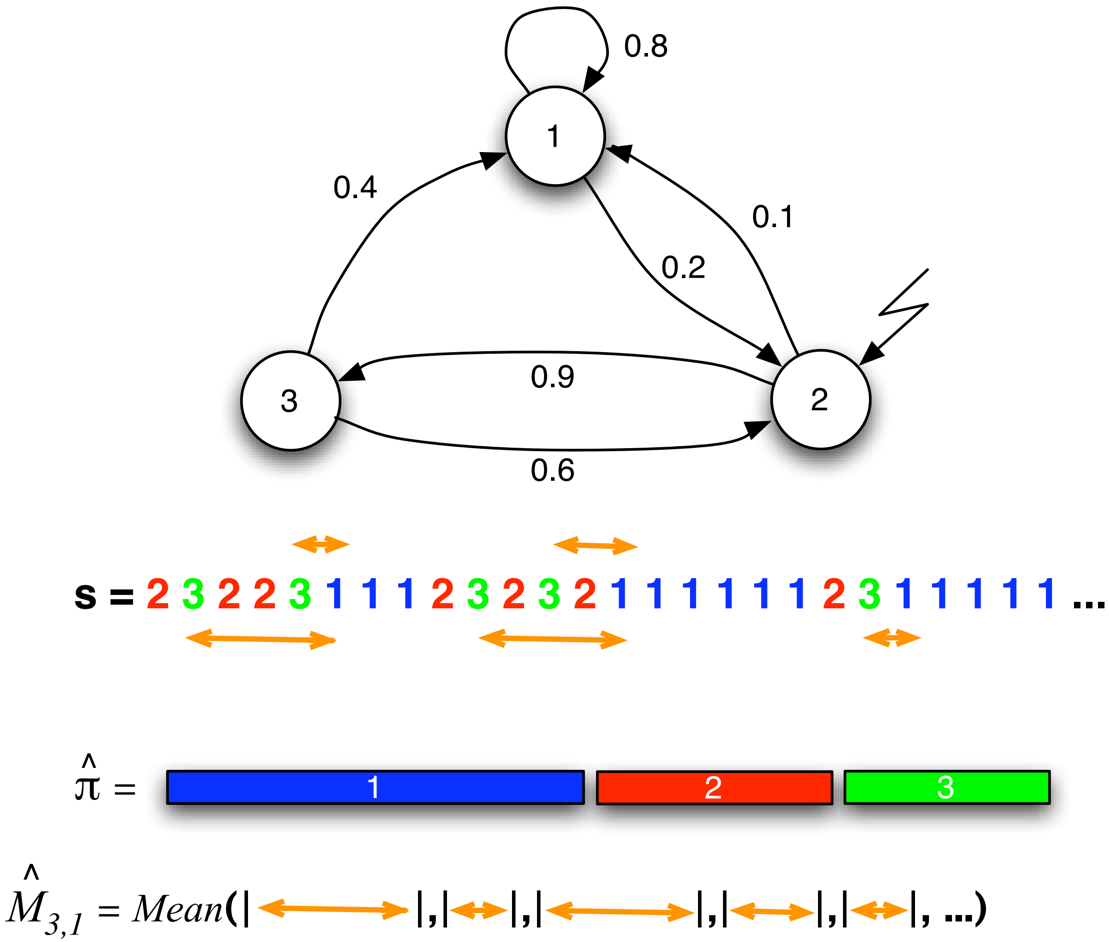
Fundamental quantities of Markov chains



s = 2 3 2 2 3 1 1 1 2 3 2 3 2 1 1 1 1 1 1 2 3 1 1 1 1 1 ...



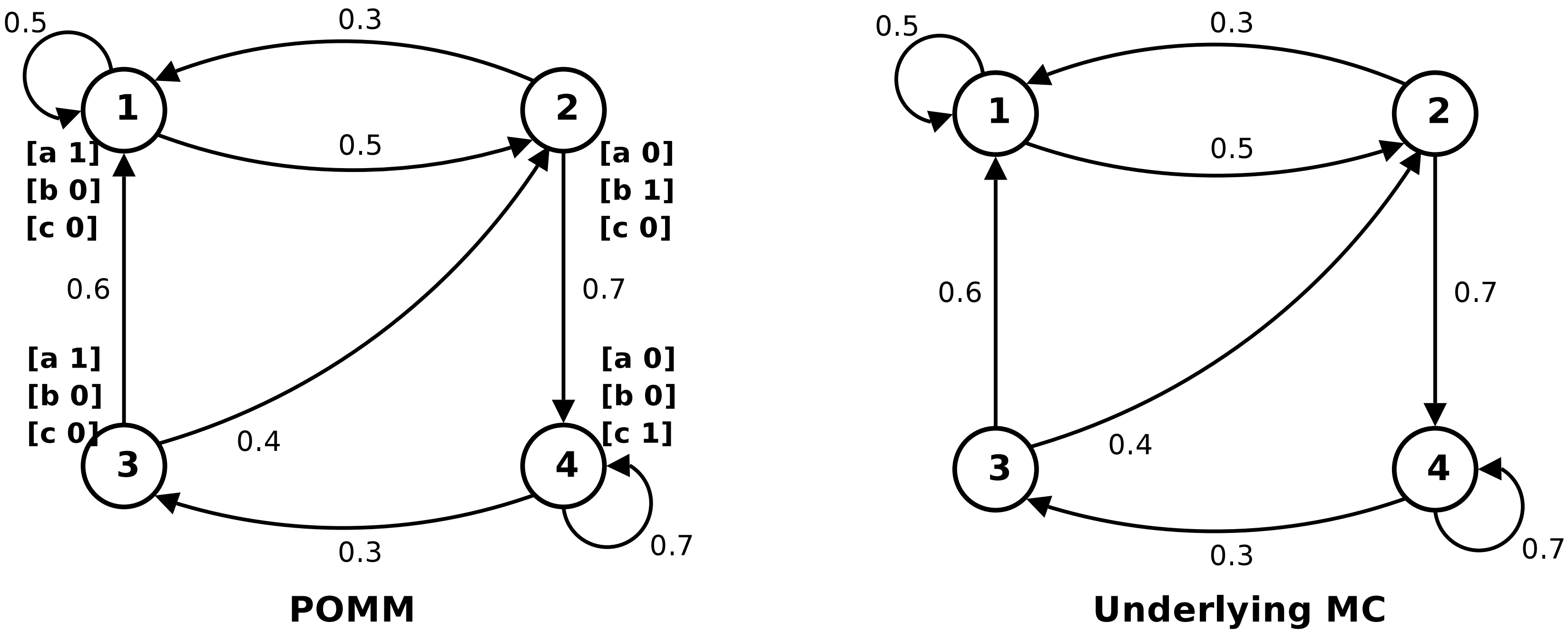
Fundamental quantities of Markov chains



The ***stationary distribution*** and the ***Mean First Passage Times (MFPT)*** can also be exactly computed using the MC transition matrix.

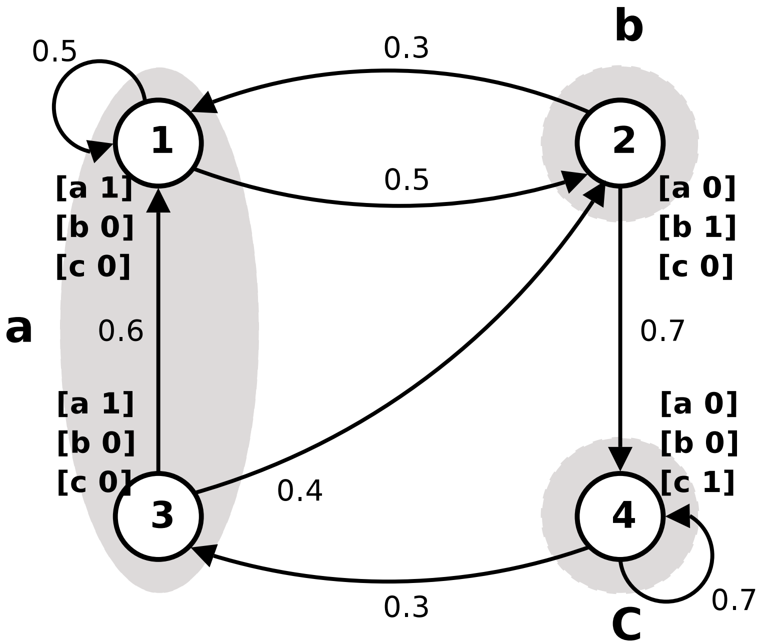
Relation between POMM and Markov chains

For any HMM, its ***underlying MC*** is obtained by ignoring the emission probabilities and the alphabet.

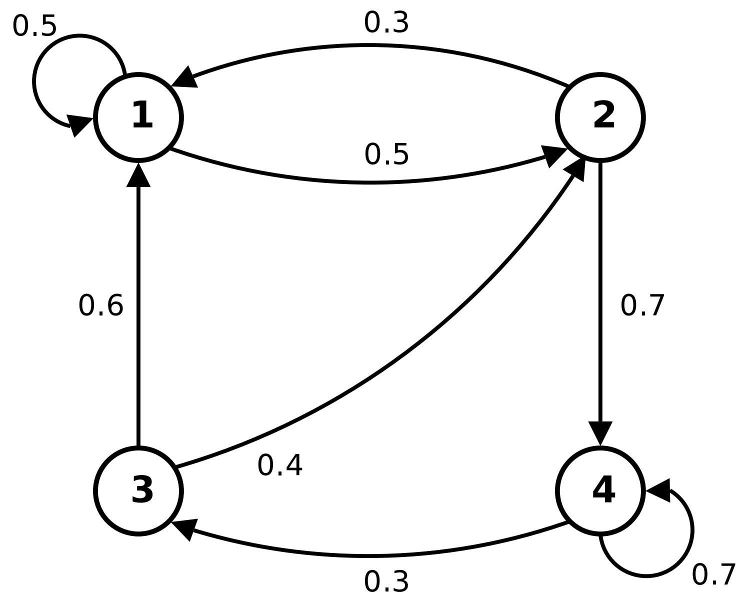


Relation between POMM and Markov chains

For any HMM, its ***underlying MC*** is obtained by ignoring the emission probabilities and the alphabet.



POMM

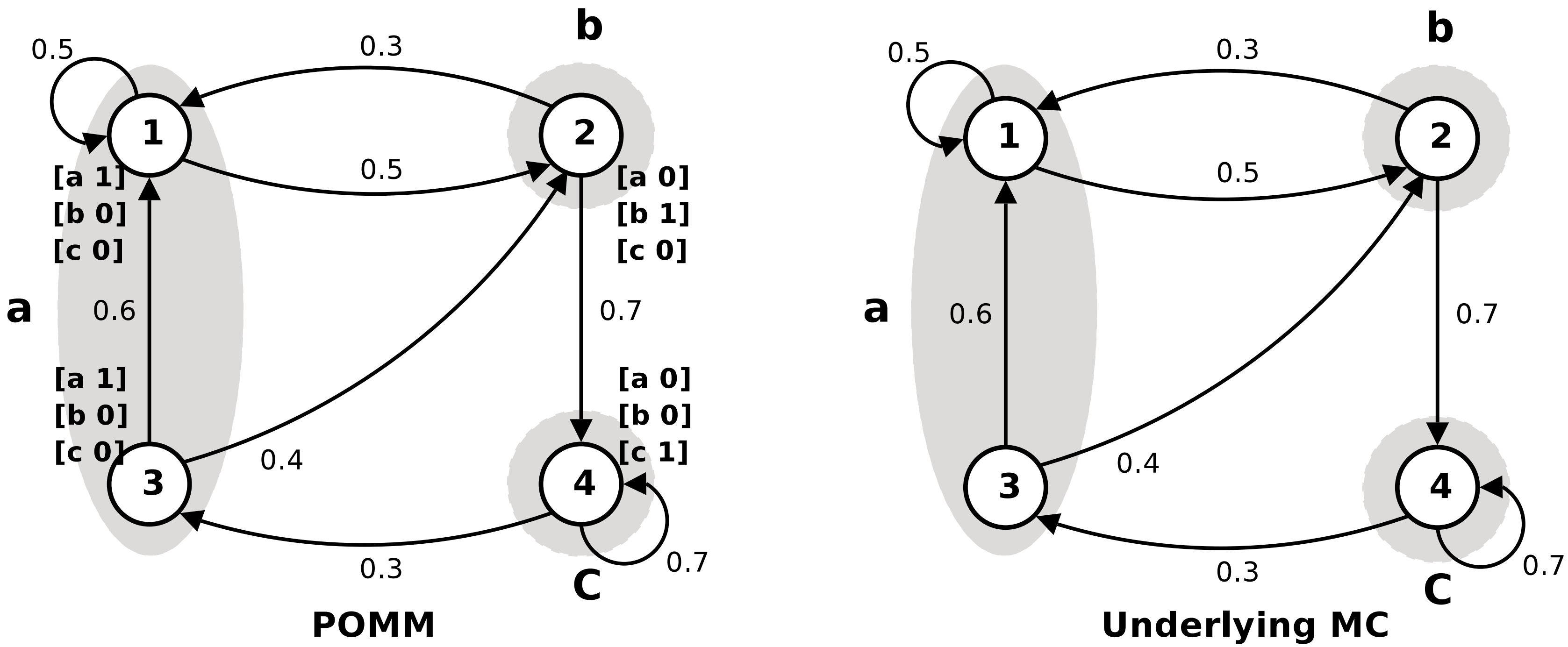


Underlying MC

The ***observable partition*** is the partition which groups all states emitting the same letter together.

Relation between POMM and Markov chains

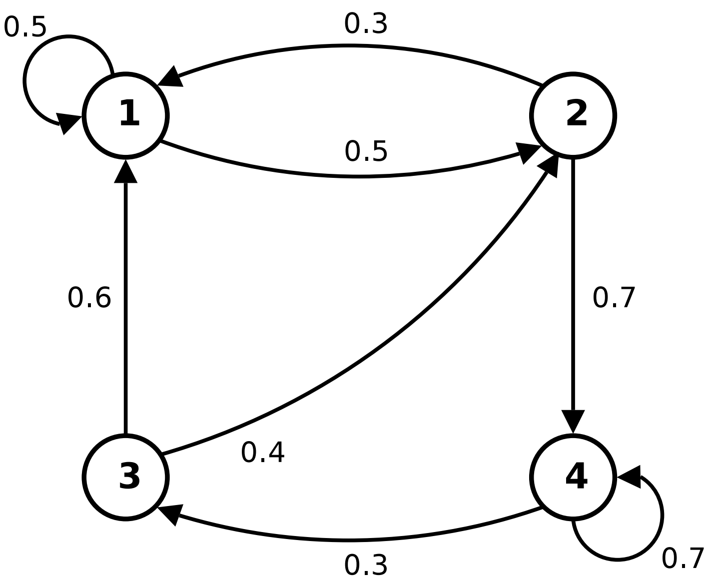
For any HMM, its ***underlying MC*** is obtained by ignoring the emission probabilities and the alphabet.



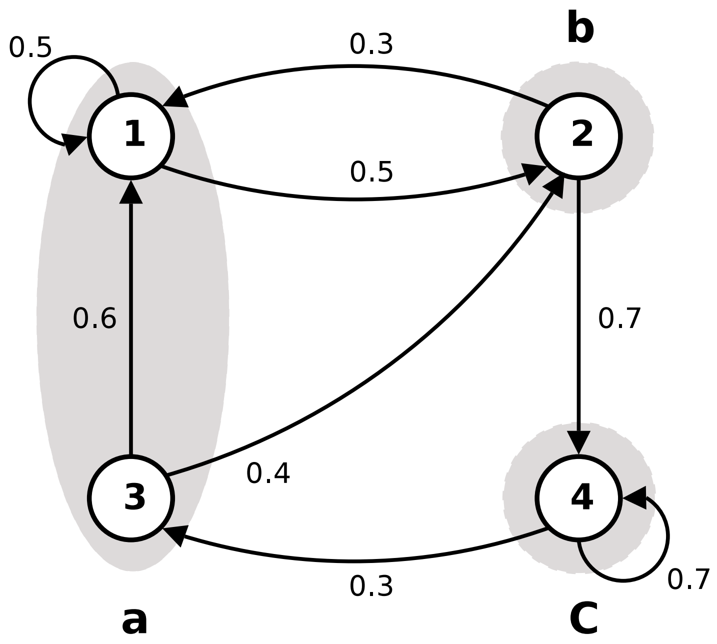
The ***observable partition*** is the partition which groups all states emitting the same letter together.

Relation between POMM and Markov chains

Lumped process with respect to a partition κ



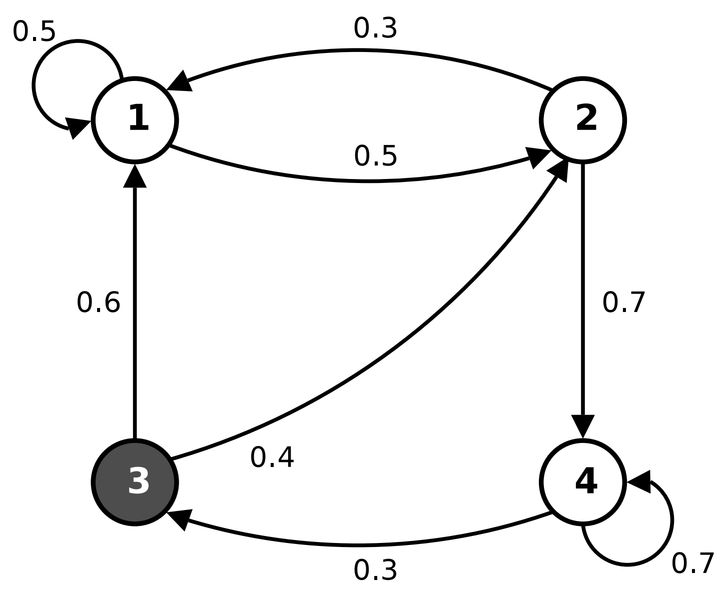
MC T



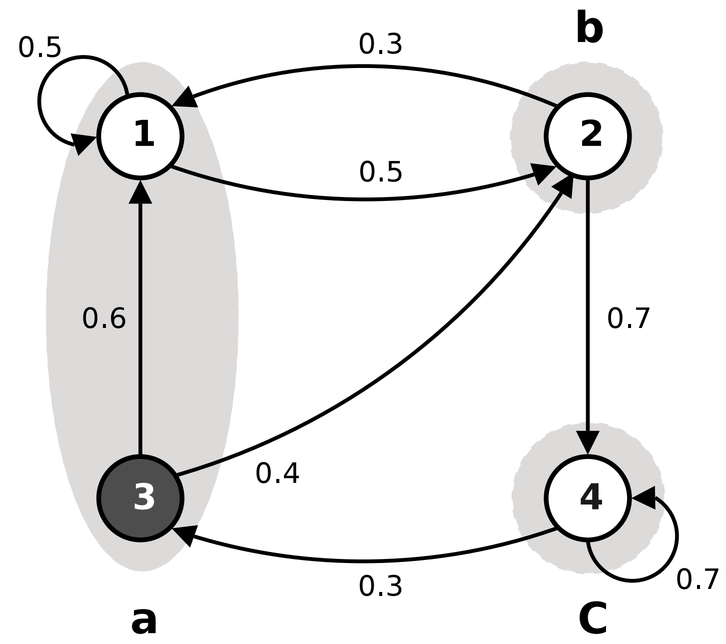
Lumped process T//k

Relation between POMM and Markov chains

Lumped process with respect to a partition κ



MC T

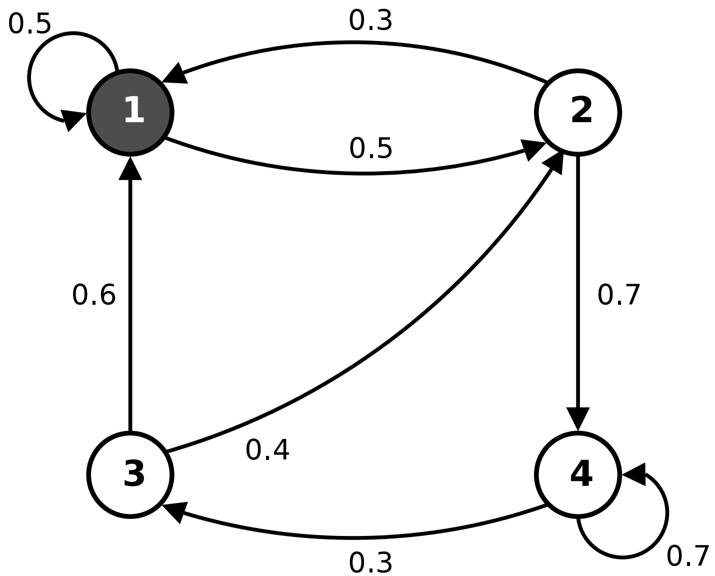


Lumped process T//k

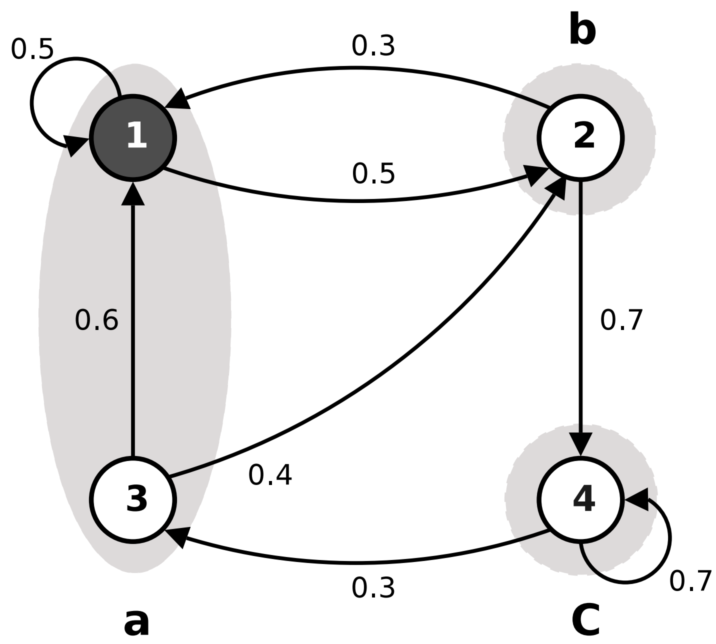
rand. walk in T	3
rand. walk in $T//\kappa$	a

Relation between POMM and Markov chains

Lumped process with respect to a partition κ



MC T

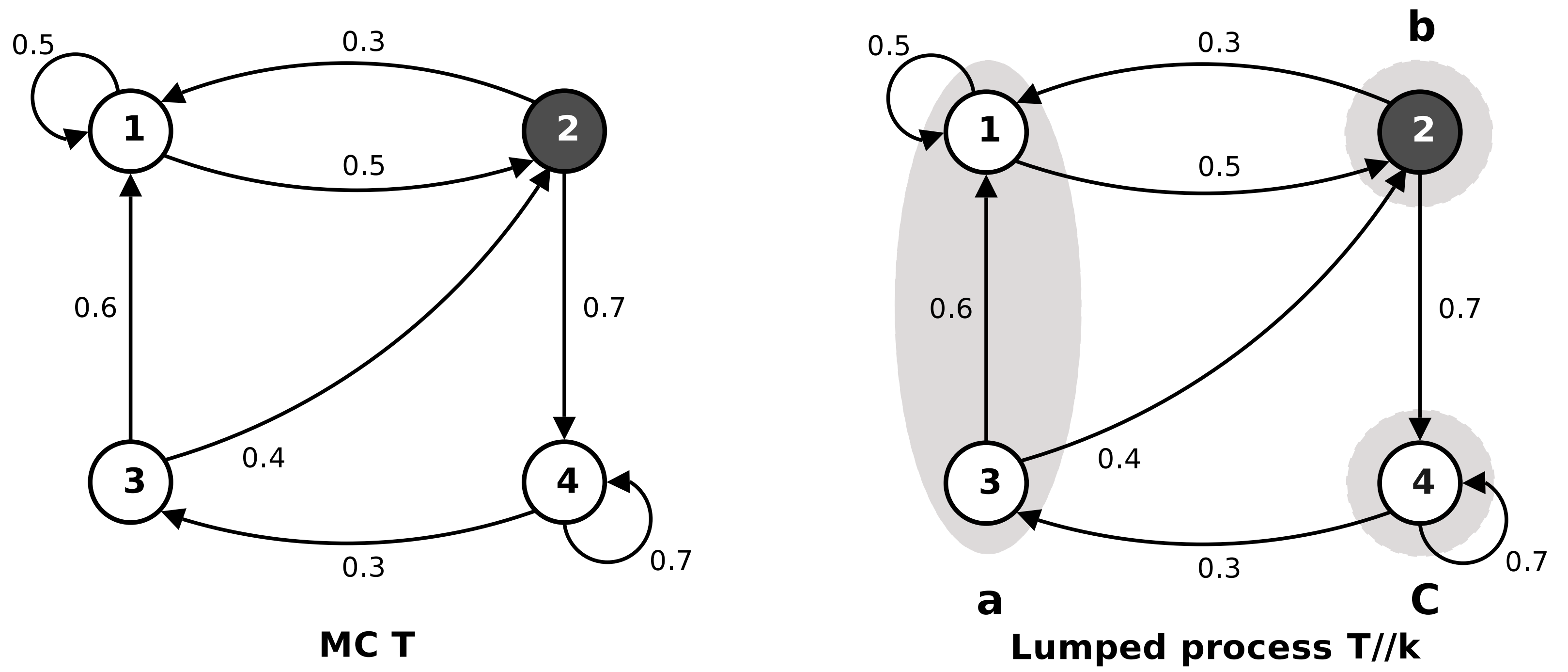


Lumped process T//k

rand. walk in T	3	1
rand. walk in $T//\kappa$	a	a

Relation between POMM and Markov chains

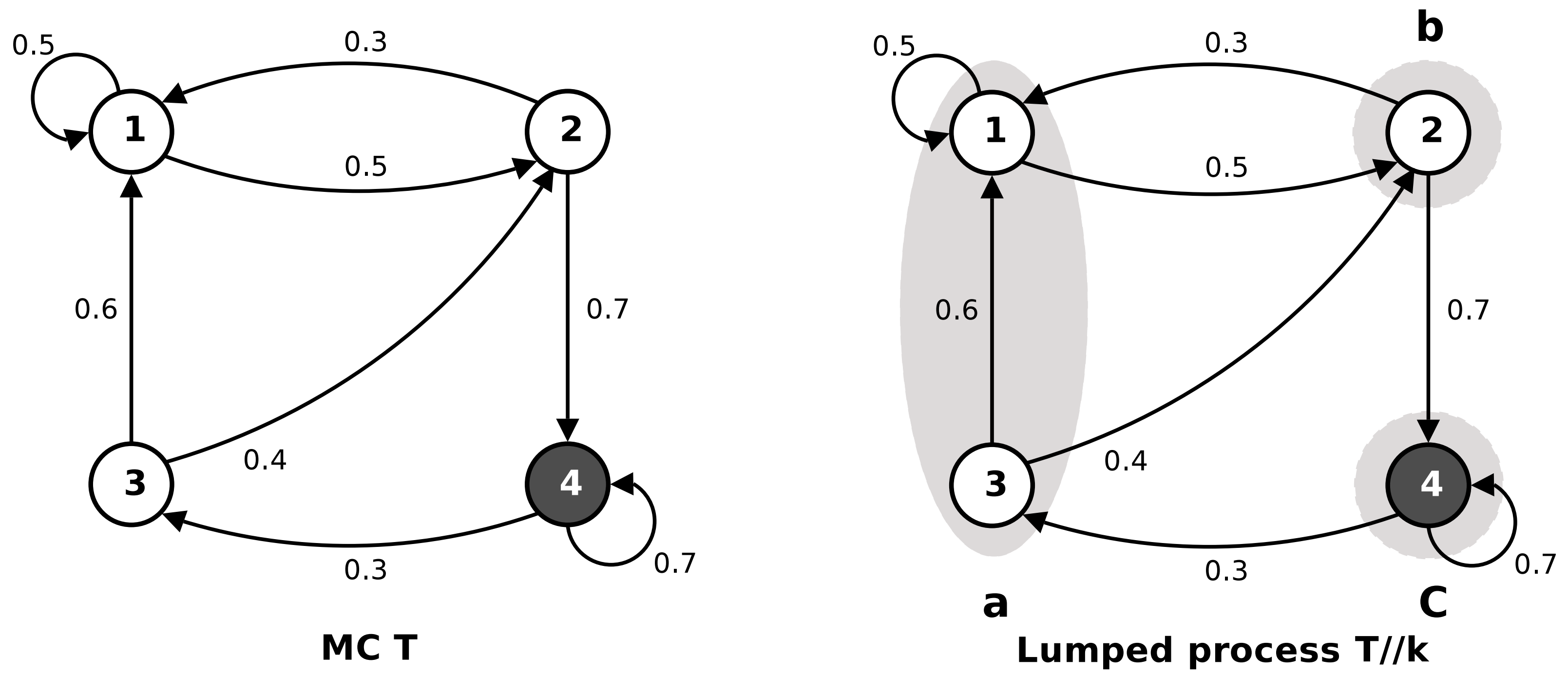
Lumped process with respect to a partition κ



rand. walk in T	3	1	2
rand. walk in $T//\kappa$	a	a	b

Relation between POMM and Markov chains

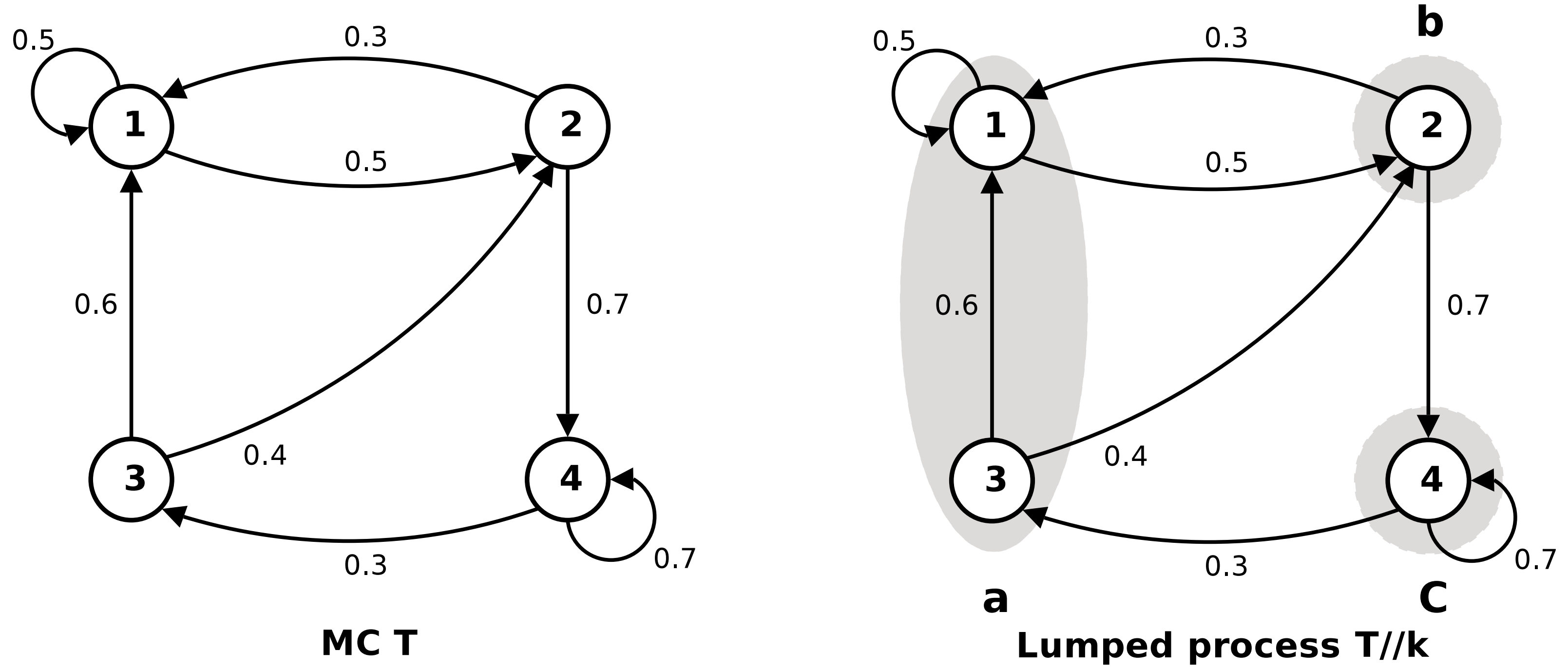
Lumped process with respect to a partition κ



rand. walk in T	3	1	2	4
rand. walk in $T//\kappa$	a	a	b	c

Relation between POMM and Markov chains

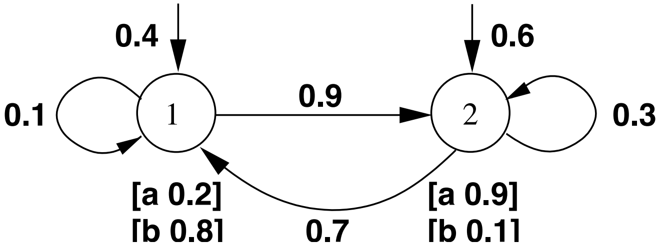
Lumped process with respect to a partition κ



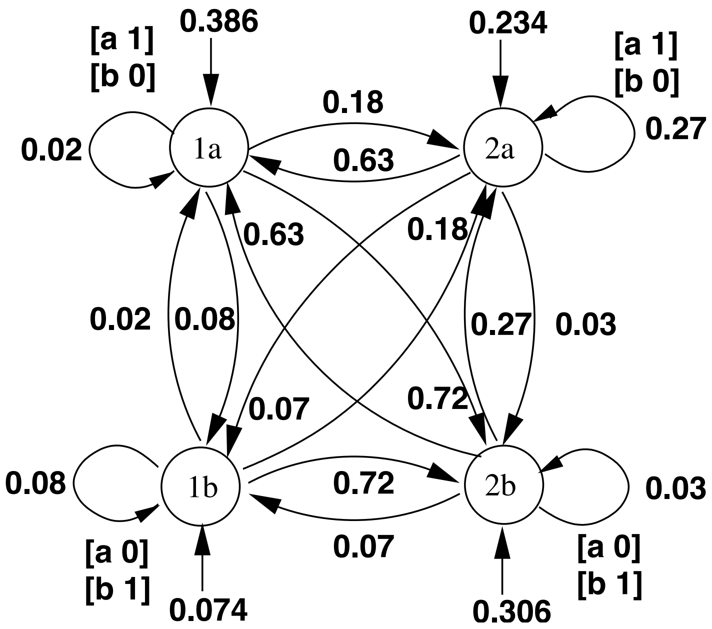
rand. walk in T	3	1	2	4	4	3
rand. walk in $T//\kappa$	a	a	b	c	c	a

A ***POMM*** can be seen as ***lumped process*** i.e. its underlying MC lumped w.r.t the *observable partition*.

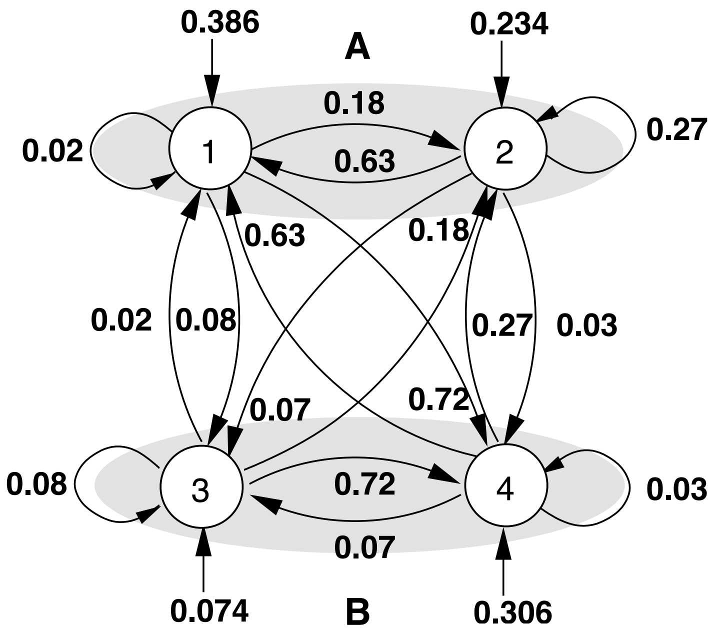
Summary



HMM



POMM



Lumped process

Random walks in HMMs $\Longleftrightarrow$ ***Random walks in lumped processes***

We are interested in the ***dynamics of random walks*** :

- Stationary distribution
- Mean First Passage Times

Our requirements

Given a string s drawn from a target POMM, we propose that the ***induced model*** has to meet the following ***requirements***:

1. Stationary distribution of the blocks has to match the letter frequencies estimated from s

2. MFPT between blocks have to match the MFPT (between the letters) extracted from s

The induced model has to fit the target dynamics estimated from the sample s

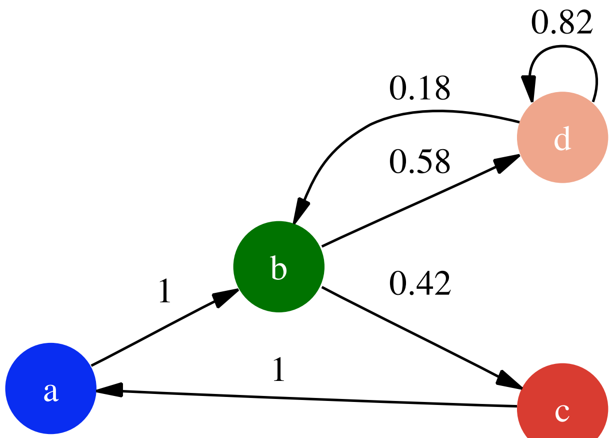
Induction algorithm

Input: -A string s from the target TP
-A precision parameter ϵ

Output: -A POMM EP

1. Initialization

- $\hat{M}$ ←Estimate MFPTs from the sample s
- EP ← Estimate ***Maximum likelihood*** model from s

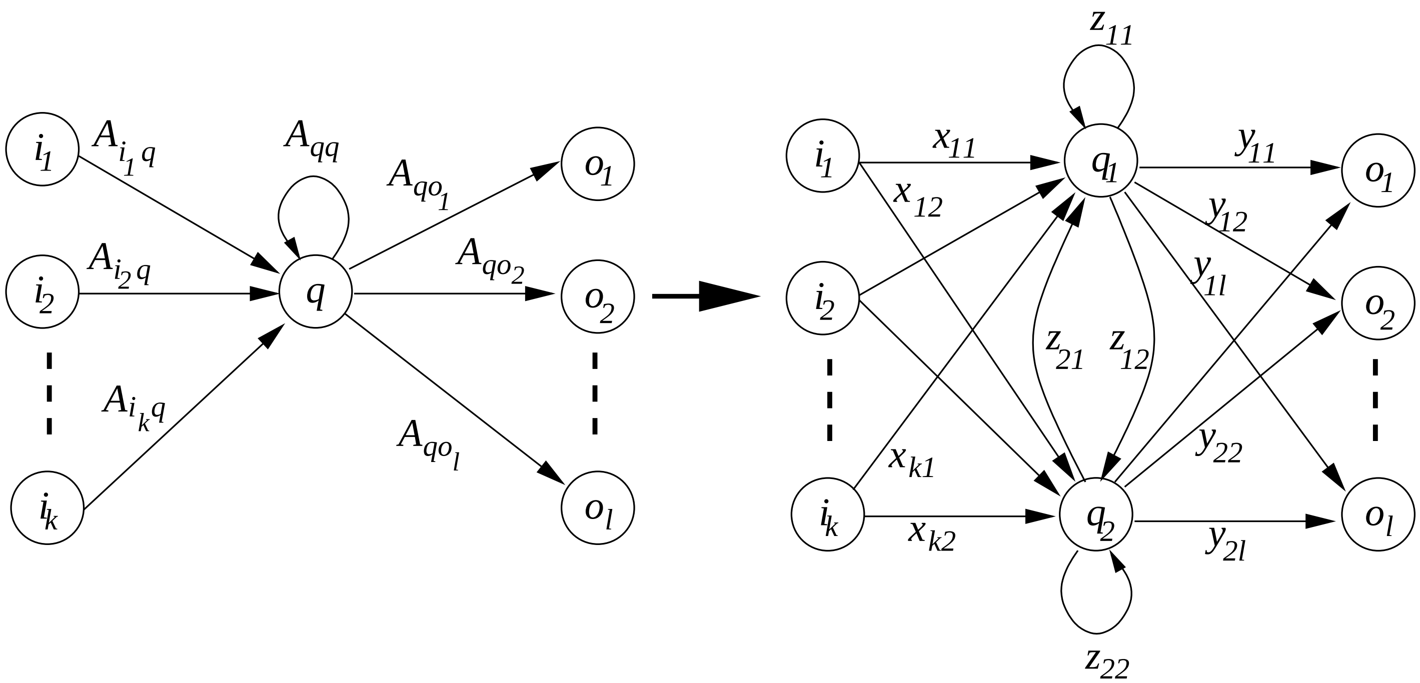


- $s = dbcabddddddbcabddb\dots$
- $A_{a,b} = \frac{count(a,b)}{count(a)}$
- Model size = $|\Sigma|$
- Satisfies stationary distribution
- ***Does not satisfy MFPT between letters***

Induction algorithm

2. Iterate

- $q \leftarrow$ Find best state to split



- Minimize $\quad W(X,Y,Z) = \sum_{i,j=1}^{|\Sigma|} (\hat{M}_{ij} - M_{ij}//\kappa)^2$

  such that $\begin{cases} \textit{Blocks stationary distribution is unchanged} \\ \textit{The model remains a proper POMM} \end{cases}$

Induction algorithm

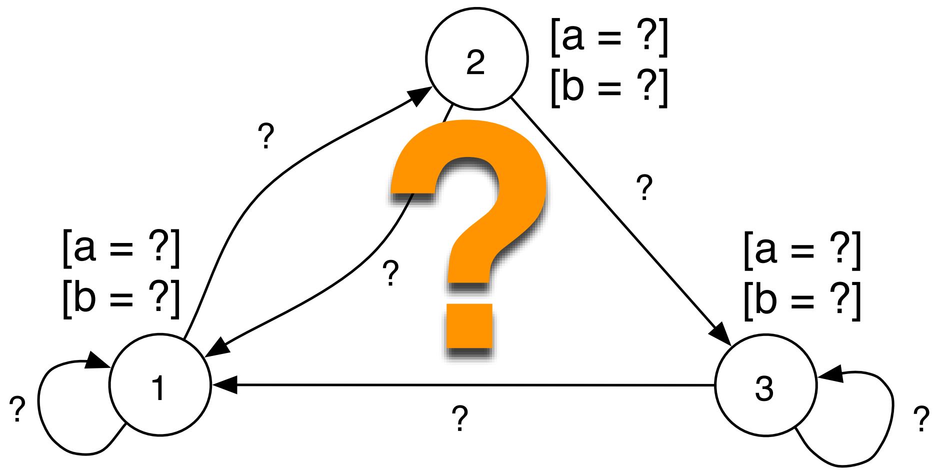
3. Termination criterion

- The value of the objective function falls below a user-defined threshold:

$$W(X, Y, Z) < \epsilon$$

Example

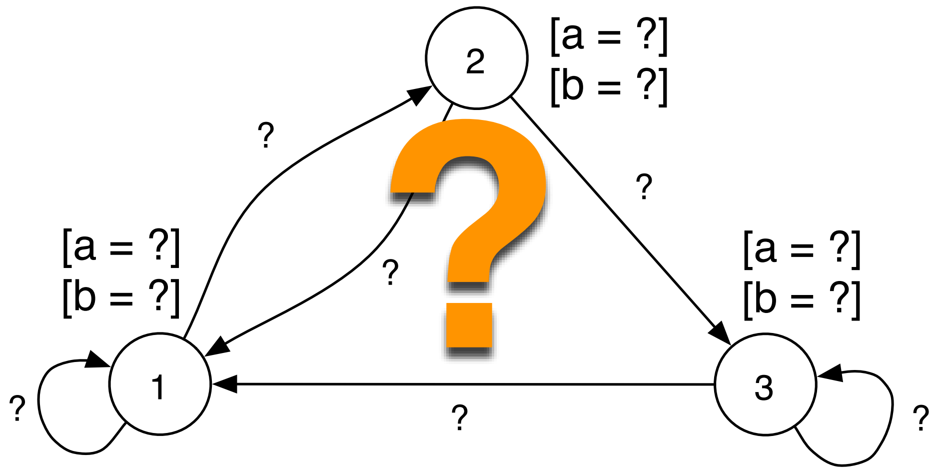
Unknown HMM target



s = b a a b a a b a a a b b b a a b b a a b ...

Example

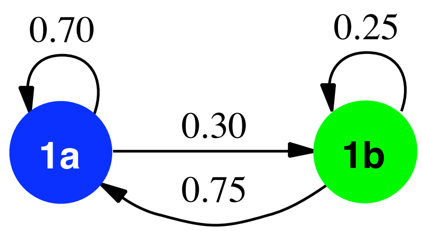
Unknown HMM target



s = b a a b a a b a a a b b b a a b b a a b ...

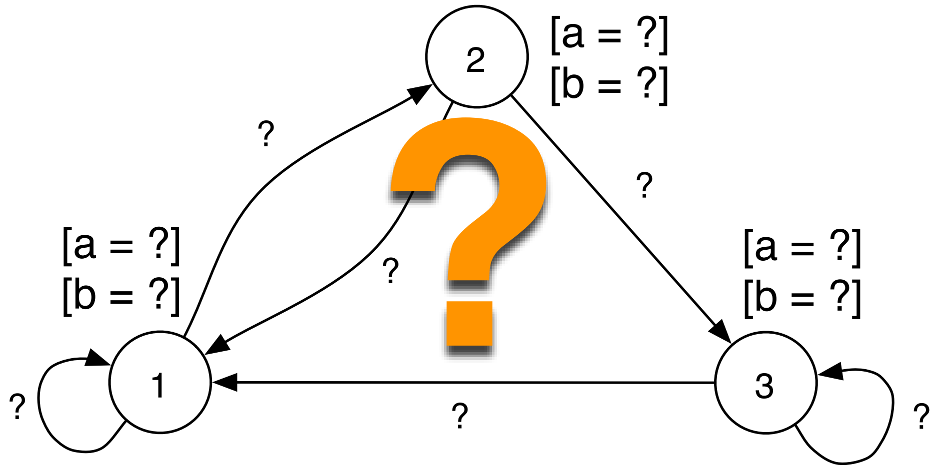
Maximum likelihood model

$$W = 3.04\ .10^{-2}$$



Example

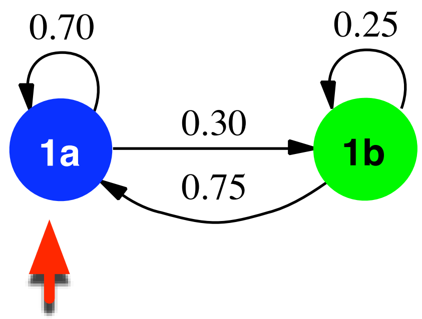
Unknown HMM target



s = b a a b a a b a a a b b b a a b b a a b ...

Maximum likelihood model

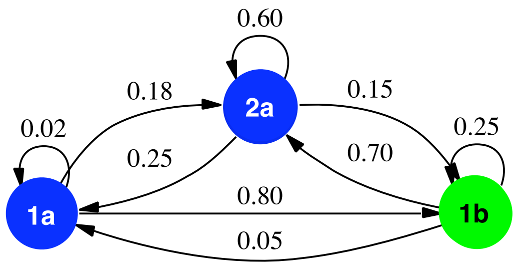
$$W = 3.04 \; .10^{-2}$$



Example

Split n°1

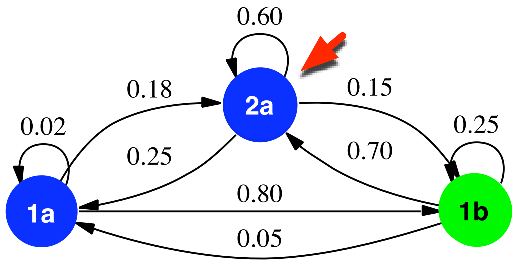
$$W = 9.03\ .10^{-3}$$



Example

Split n°1

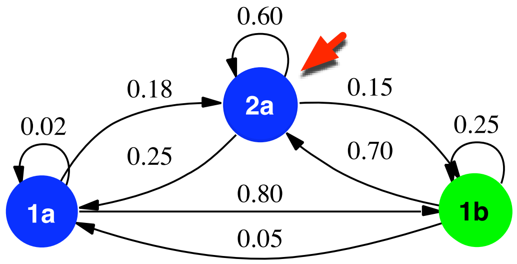
$$W = 9.03\ .10^{-3}$$



Example

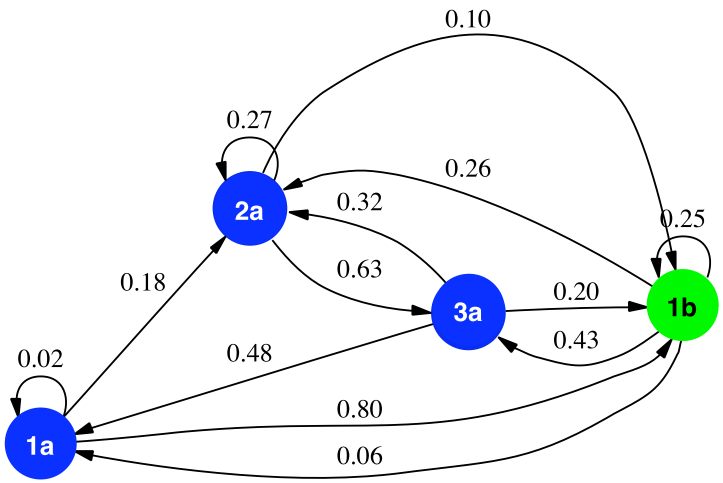
Split n°1

$$W = 9.03\ .10^{-3}$$



Split n°2

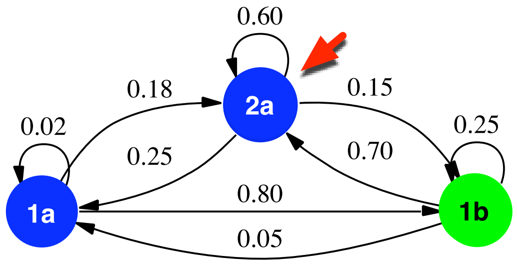
$$W = 2.93\ .10^{-4}$$



Example

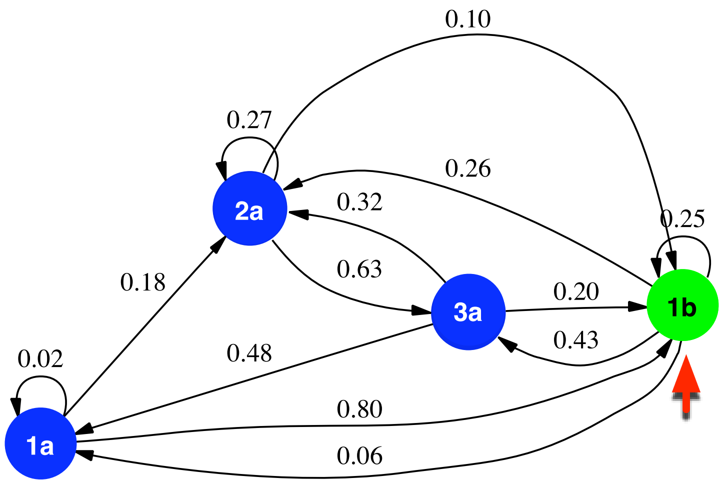
Split n°1

$$W = 9.03\ .10^{-3}$$



Split n°2

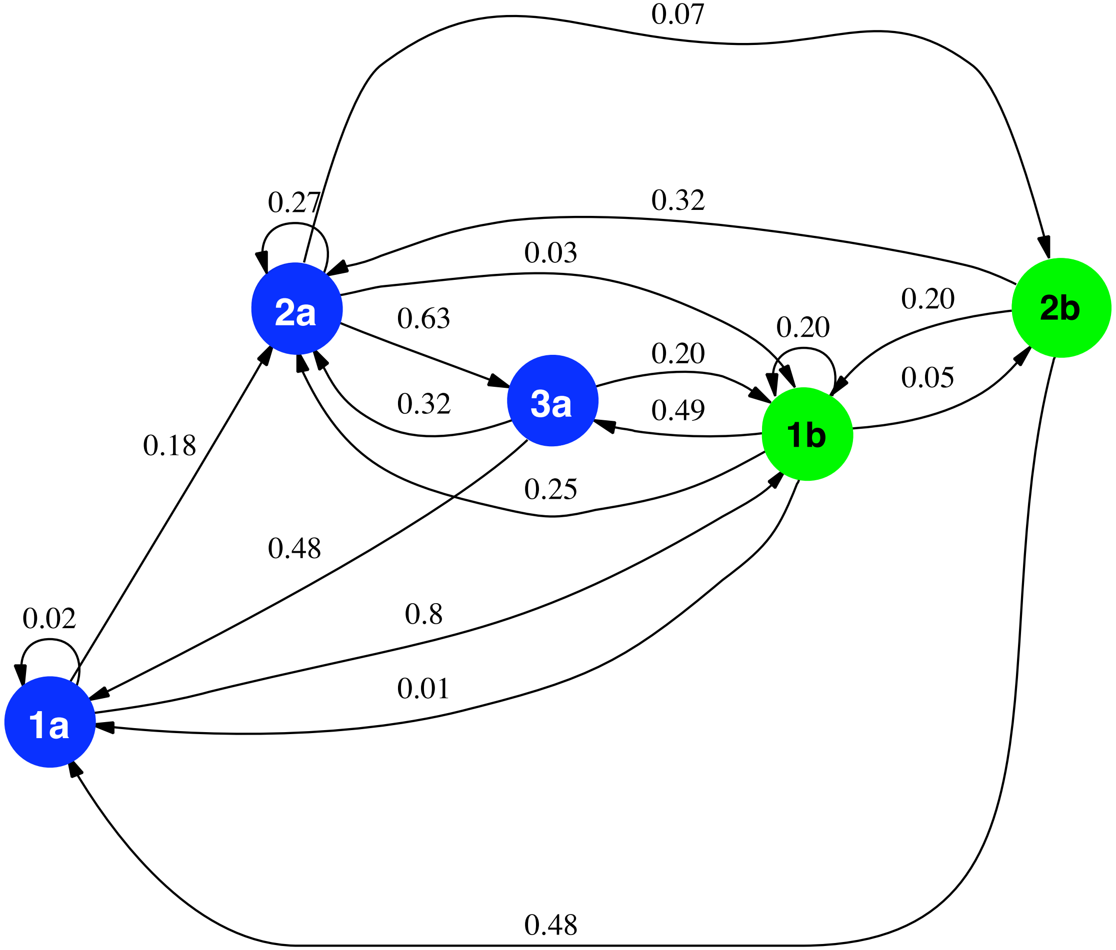
$$W = 2.93\ .10^{-4}$$



Example

Split n°3

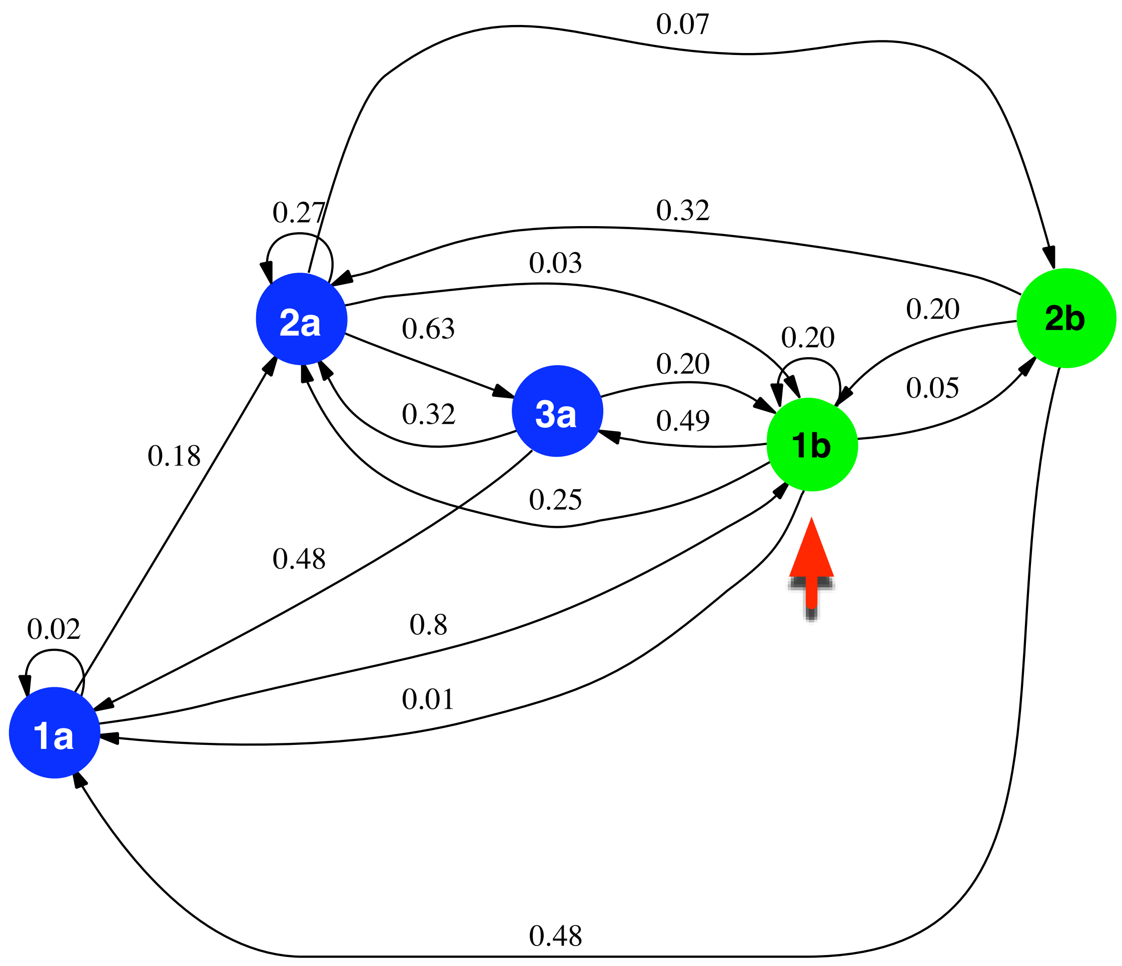
$W = 2.32\ .10^{-4}$



Example

Split n°3

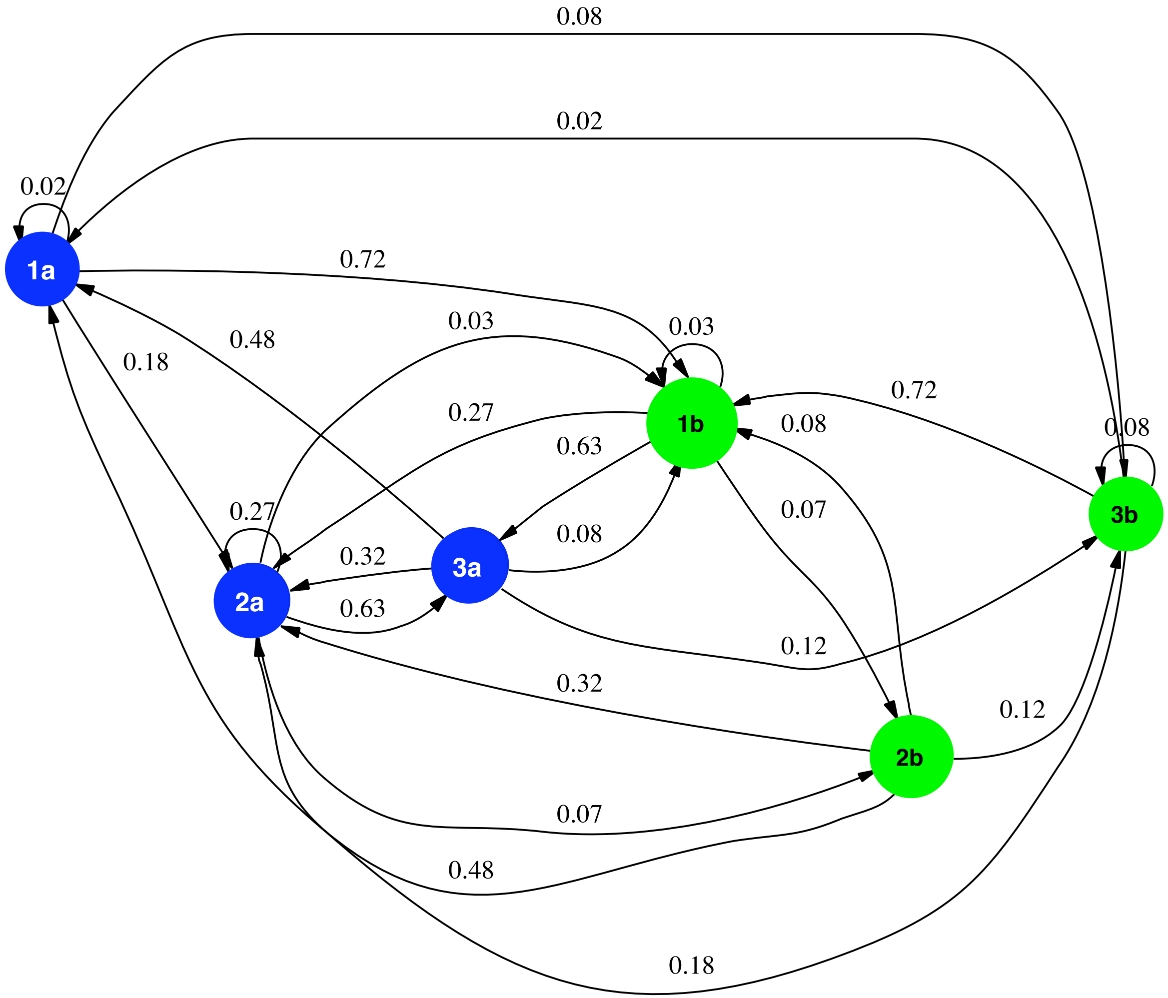
$$W = 2.32\ .10^{-4}$$



Example

Split n°4

$$W = 1.00\ .10^{-5}$$



Conclusion and future work

- We proposed a novel approach to induce HMMs based on the observed dynamics in the sample.

Ongoing work

- Systematic experimental study of the method
  To do so, we are working on:
  - A better criterion to find the best state to split
  - An efficient procedure to solve the optimization problem

Future work

- Formal study of convergence
- Relations with the EM algorithm