Hes·so
Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences
Western Switzerland

MSE | MASTER OF SCIENCE IN ENGINEERING

Multimodal Processing, Recognition and Interaction

Introduction HMM

Elena Mugellini, Jean Hennebert, Stefano Carrino

Modèles de Markov Cachés (HMMs) Partie 1

- Evaluation des performances – intervalles de confiance
- Introduction aux HMMs
  - Les chaînes de Markov à temps discret
  - Extension aux HMMs
  - Probabilités d'émission
  - Les éléments d'un HMM
  - Topologies des HMMs
- Problèmes résolus par les HMMs

Evaluation des performances

INTERVALLES DE CONFIANCE

Evaluation des performances – intervalles de confiance

- Donner un résultat sans indication sur sa précision n'a que peu d'intérêt car il n'est pas reproductible.
- Exemple: lancer de pièce truquée, pour laquelle la probabilité P d'observer pile est inconnue. La fréquence empirique d'observer pile est l'estimateur naturel de P. Si sur 100 lancers on obtient 60 pile, l'estimation (ponctuelle) proposée pour P est 0.60. Mais ce résultat n'est pas reproductible. Si on renouvelle les 100 lancers, on obtiendra probablement des estimations différentes.

Intervalles de confiance - Motivation

- Plutôt que de donner une estimation ponctuelle, on proposera un **intervalle**, choisi de manière à contrôler par un **niveau de confiance**, les chances que le résultat aurait d'être confirmé si on renouvelait l'expérience. On cherche à distinguer les valeurs du paramètre pour lesquelles l'observation (60 pile sur 100 lancers) est plausible, des valeurs pour lesquelles elle est trop peu vraisemblable.

Intervalles de confiance - Définition

Intervalle de confiance: *l'intervalle de confiance au risque α d'une valeur que l'on cherche à estimer est un intervalle qui contient avec une probabilité 1- α la valeur cherchée.*

- Pour une probabilité estimée, l'intervalle de confiance est donné par:

$$IC_{1-\alpha} = p \pm z_\alpha \sqrt{\frac{p(1-p)}{n}}$$

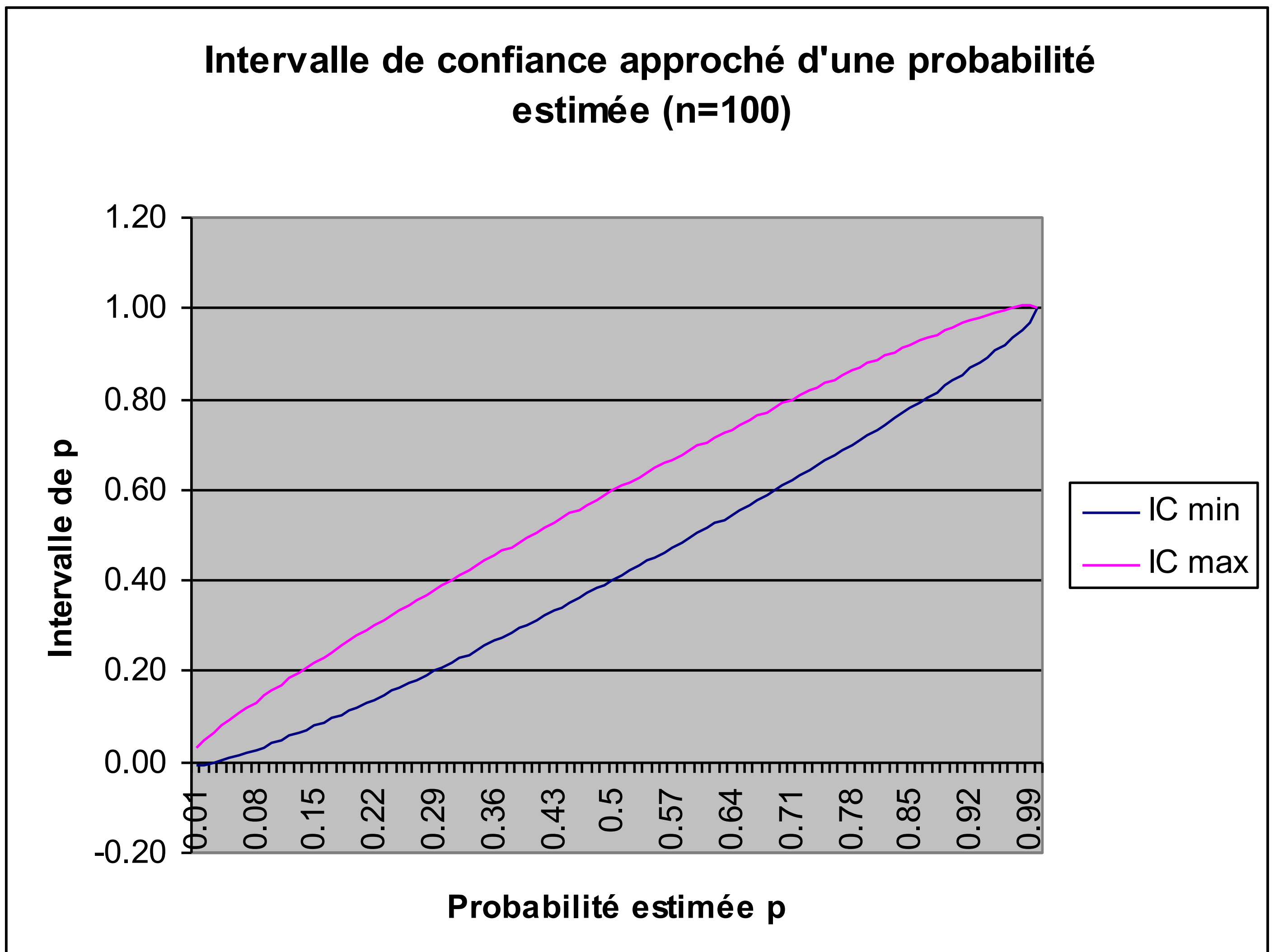
α	0.01	0.02	0.05
z_α	2.5758	2.3263	1.96

Condition de validité :

$$nIC_{\min} \geq 5,\ n(1 - IC_{\max}) \geq 5 \longrightarrow$$

Si la condition n'est pas respectée, on ne peut plus utiliser cette formule qui est pessimiste

Evaluation des performances – intervalles de confiance



- Exemple: $n = 100$, $\alpha = 0{,}05$, $p = 0{,}75$
  - $IC_{min} = 0.75 - 0.08$
  - $IC_{max} = 0.75 + 0.08$
  - Cond de validité:
    - 100 (0.75-0.08) = 67
    - 100 (1-(0.75+0.08)) = 17
- Idem mais $n = 1000$
  - $IC_{min} = 0.75 - 0.03$
  - $IC_{max} = 0.75 + 0.03$
  - Cond de validité:
    - 1000 (0.75-0.03) = 723
    - 1000 (1-(0.75+0.03)) = 223

Introduction

HIDDEN MARKOV MODELS

HMMs - Introduction

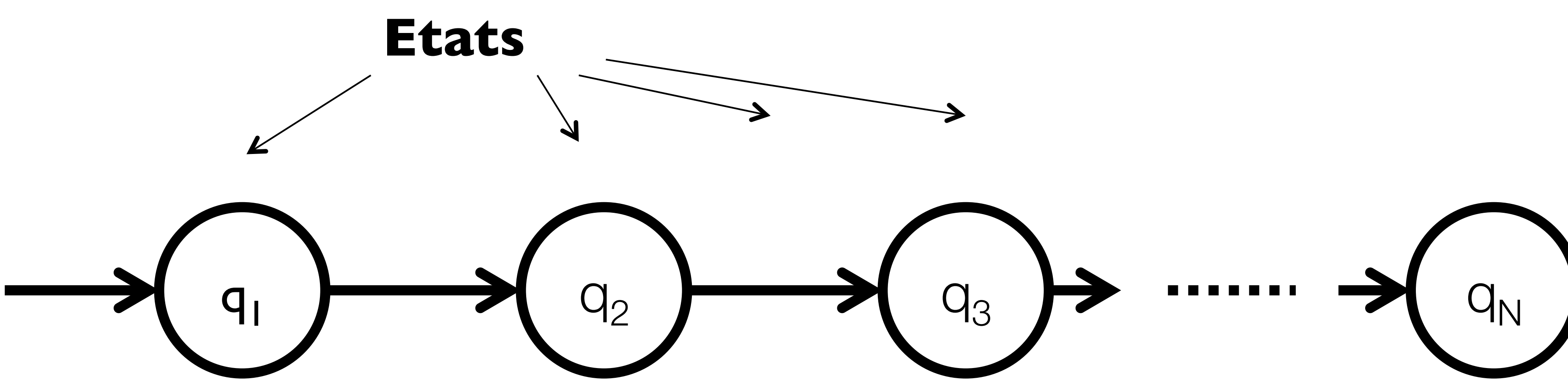
- Bases théoriques publiées par L. E. Baum dans les années 60
- Premières implémentations en traitement de la parole dans les années 70 chez IBM
- Utilisation plus récente dans d'autres domaines: reconnaissance de l'écriture, reconnaissance des gestes, biosciences
- Autres dénominations:
  - « Probabilistic functions of Markov chain »
  - « Markov sources »
- Les HMMs peuvent être vues comme une extension des chaînes de Markov

HMMs - Introduction

- Utilisés aujourd'hui dans des domaines très variés:
  - La reconnaissance automatique de la parole
  - La reconnaissance de l'écriture
  - La biométrie : vérification du locuteur et du scripteur
  - La bio-informatique : recherche dans des séquences ADN, modélisation de protéines
  - La linguistique : modélisation de texte
  - …

HMMs - Introduction

- Markov model:
  - *Le future est indépendant du passé donné le présent*



Les chaînes de Markov à temps discret

- Soit un système qui est dans un état *i* parmi un ensemble de *N* états
- Le système change d'état à chaque temps discret en fonction d'un ensemble de probabilités de transition associées à chaque état
- Processus stochastique markovien d'ordre 1: processus sans mémoire, càd:

$$P(q_t = j \mid q_{t-1} = i, q_{t-2} = k, \ldots) = P(q_t = j \mid q_{t-1} = i) = a_{ij}$$

avec

$$\sum_{j=1}^{N} a_{ij} = 1$$

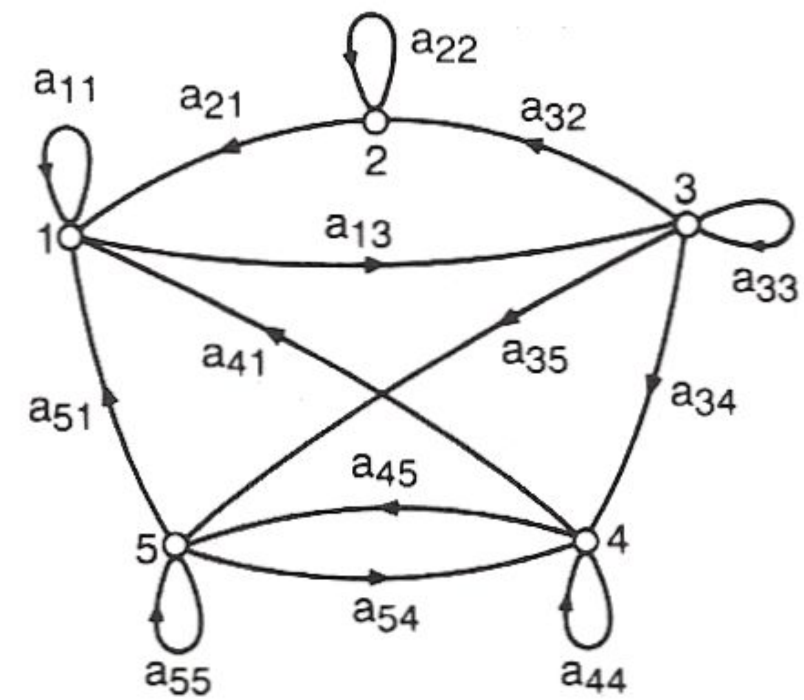


Figure 6.1 A Markov chain with five states (labeled 1 to 5) with selected state transitions.

Les chaînes de Markov à temps discret

- Exemple: modèle de Markov du temps mesuré de jour en jour sur une année
  - État 1 = précipitations
  - État 2 = nuageux
  - État 3 = ensoleillé

- Quelle est la probabilité d'avoir une suite de jours: soleil-soleil-soleil-pluie-pluie-nuageux-soleil?

Remarque: en plus des probabilités de transition, le modèle peut aussi définir les probabilités d'état initial π_i (définit la probabilité de commencer dans l'état i)

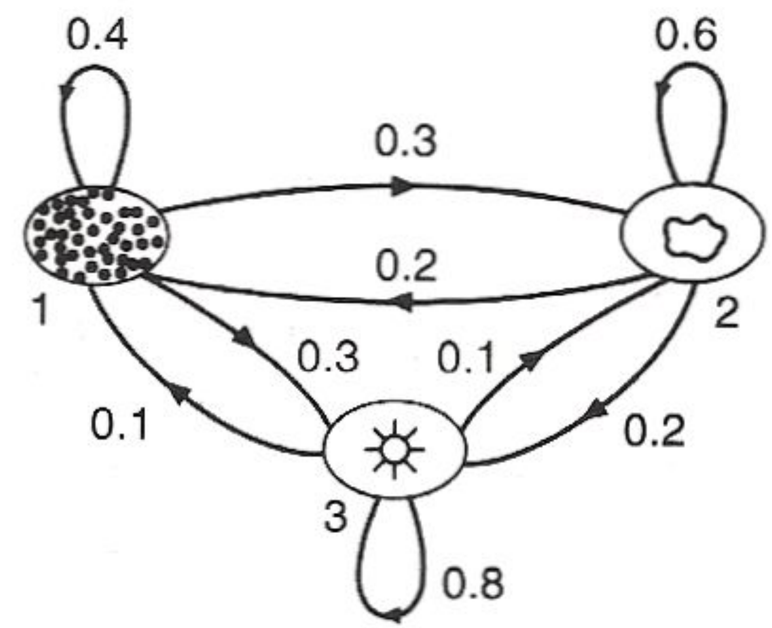
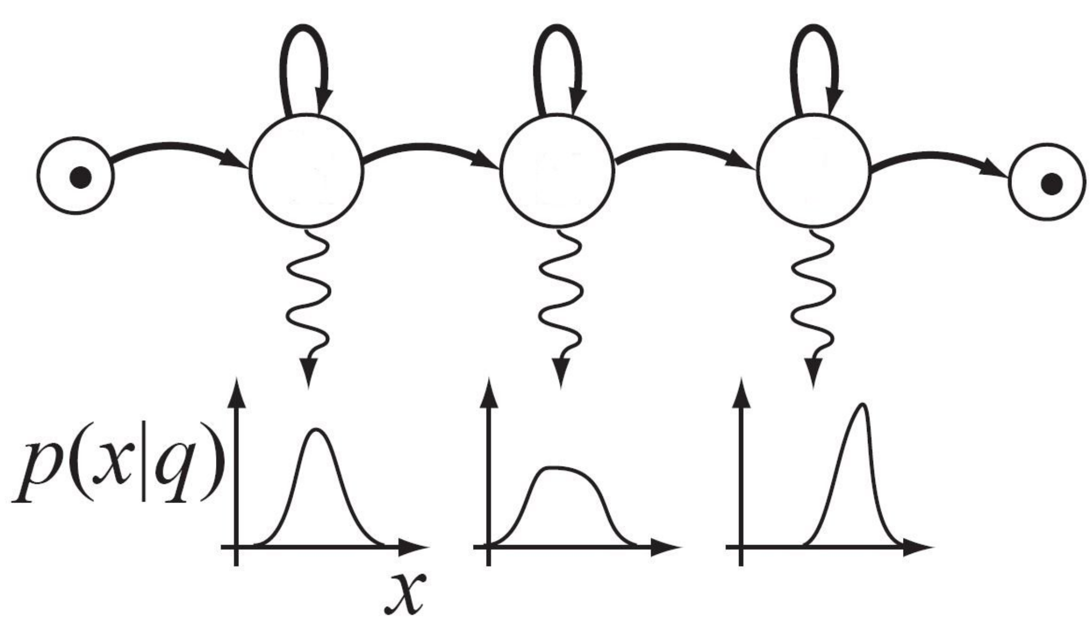


Figure 6.2 Markov model of the weather.

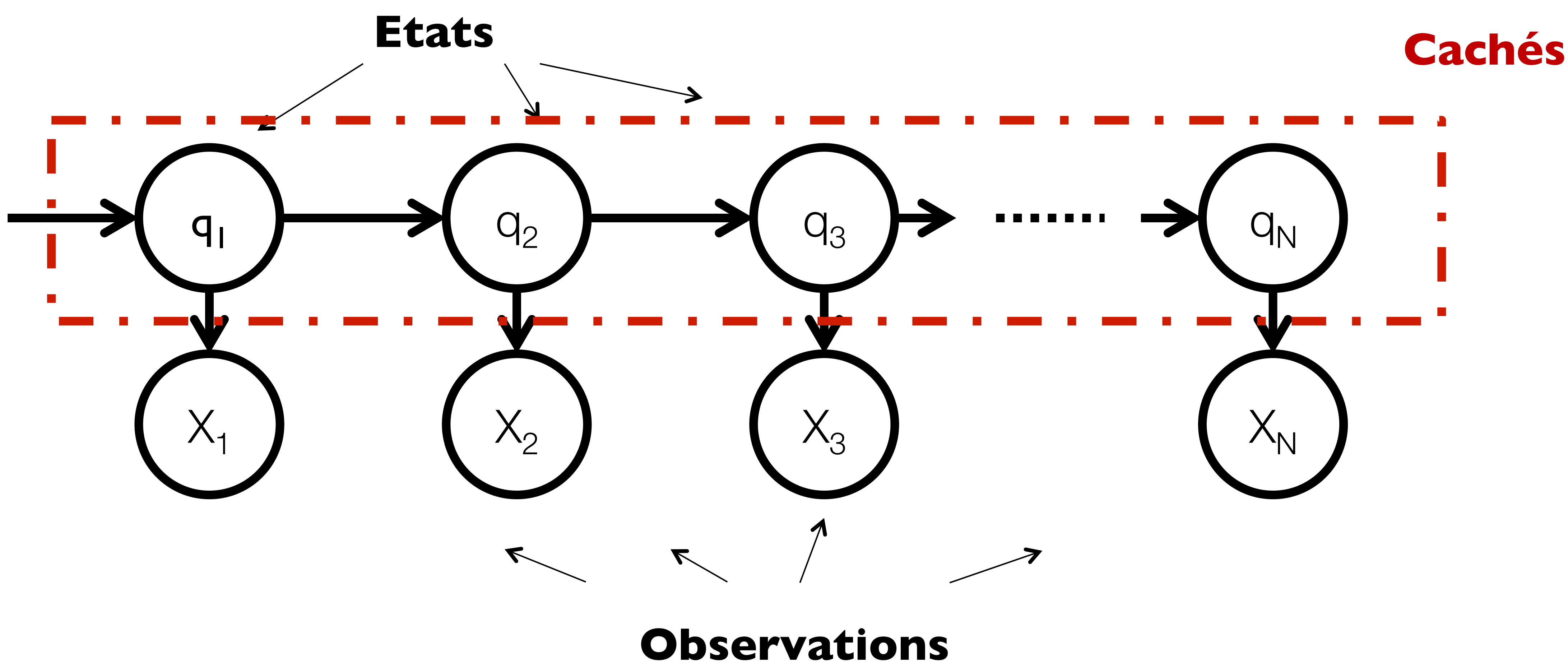
$$
\begin{aligned}
P(X \mid Modèle) &= P(s,s,s,p,p,n,s \mid Modèle) \\
&= P(s \mid s)P(s \mid s)P(p \mid s)P(p \mid p)P(n \mid p)P(s \mid n) \\
&= a_{33}a_{33}a_{31}a_{11}a_{12}a_{23} \\
&= (0.8)^2(0.1)(0.4)(0.3)(0.2) \\
&= 1.536 \times 10^{-3}
\end{aligned}
$$

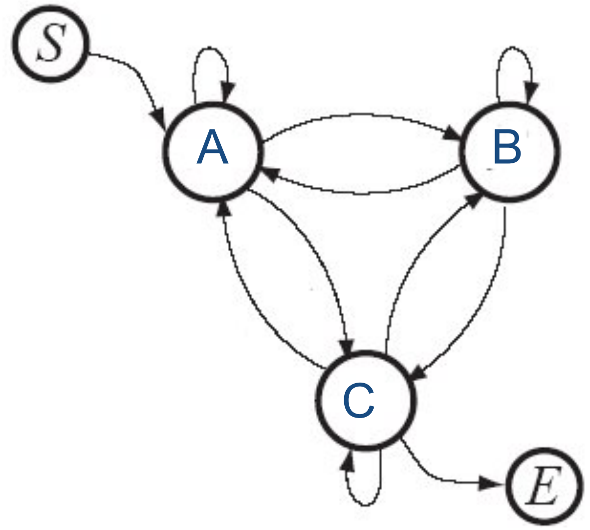
Extension aux modèles de Markov cachés

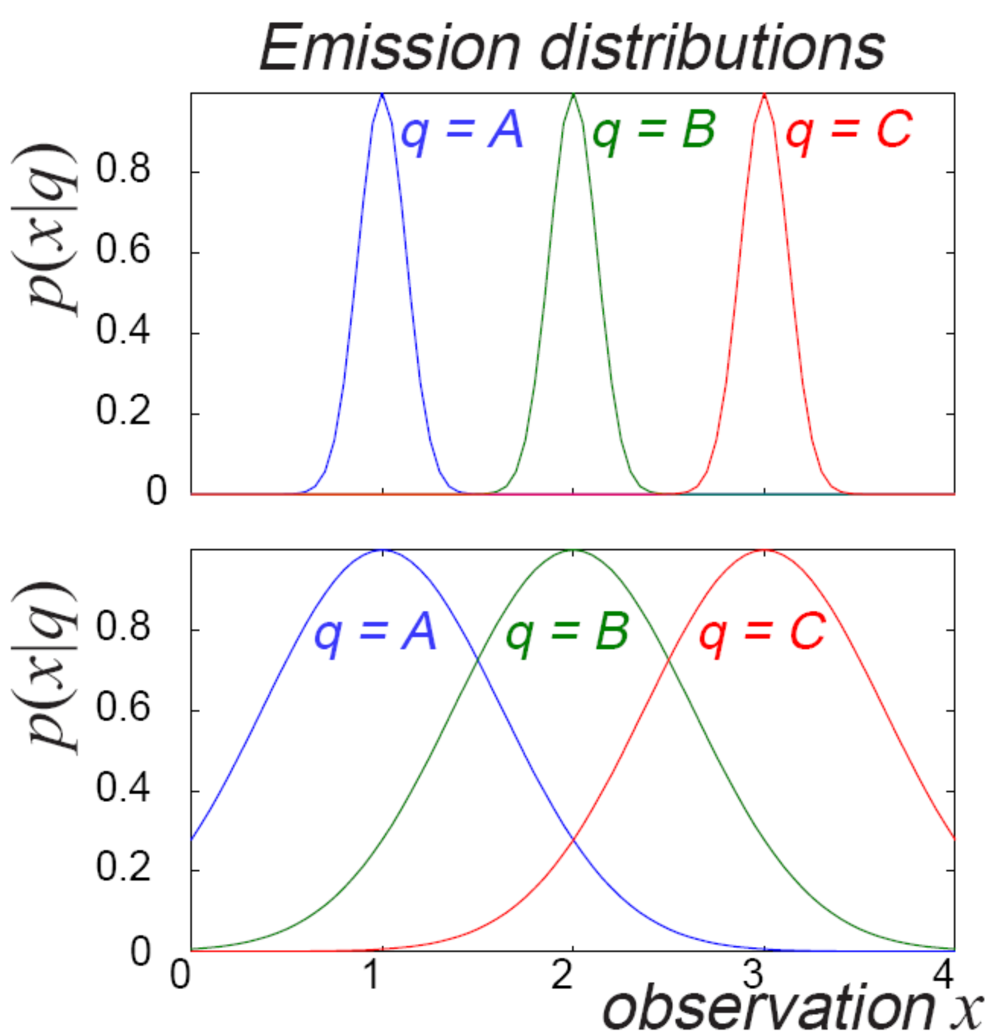
- L'observation devient une fonction probabiliste de l'état.
- Autrement dit, l'observation x a une certaine probabilité d'être « émise » dans un état q
- On parle de probabilités d'émission $p(x/q)$

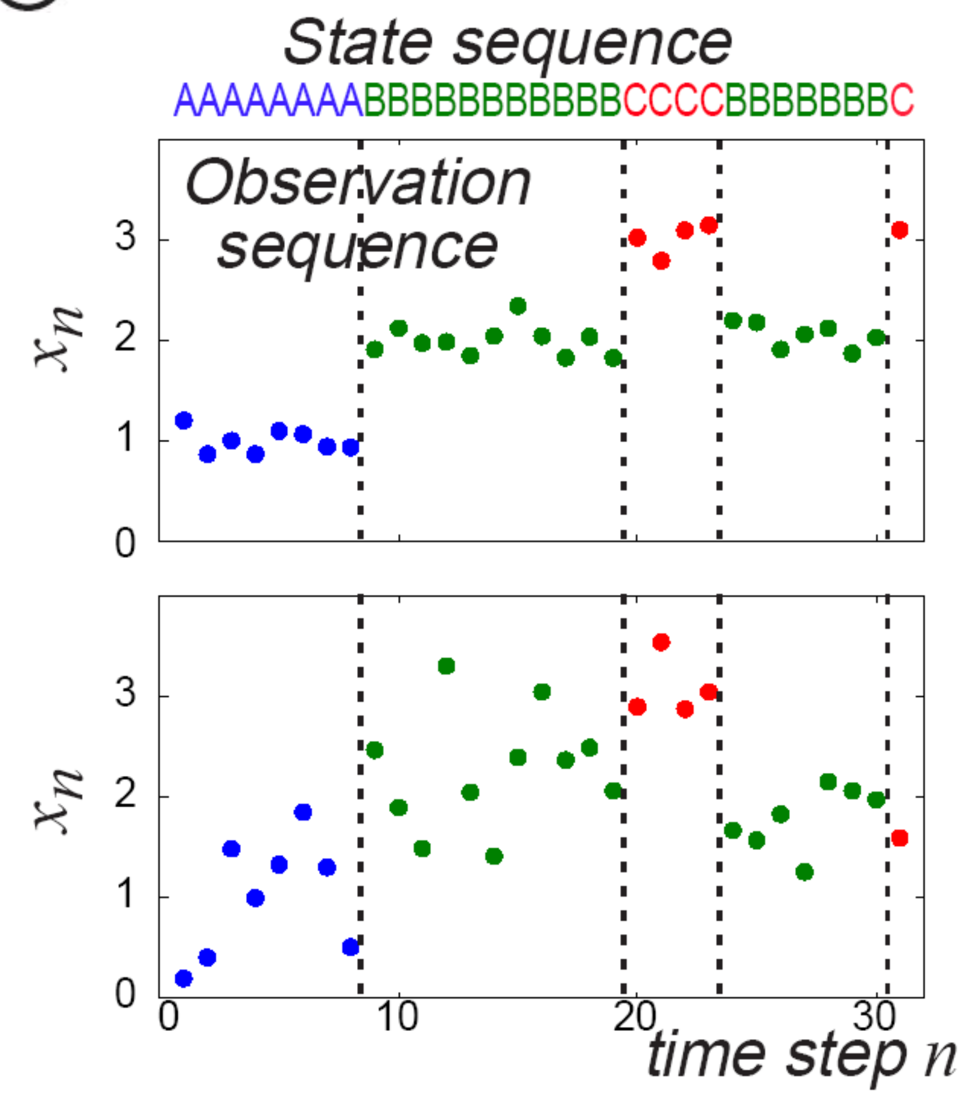


Extension aux modèles de Markov cachés

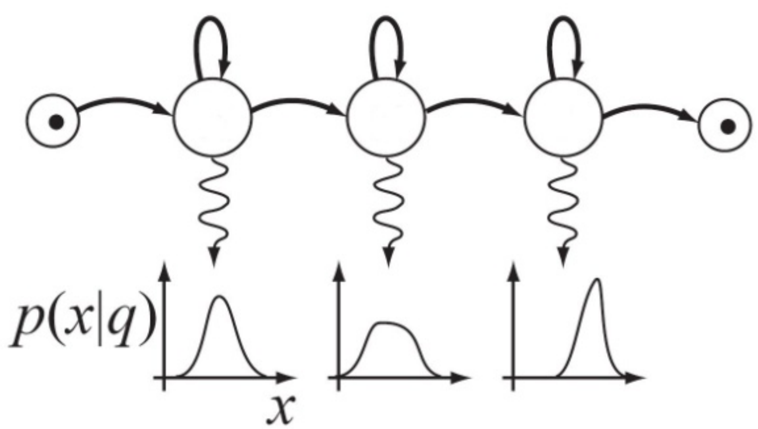


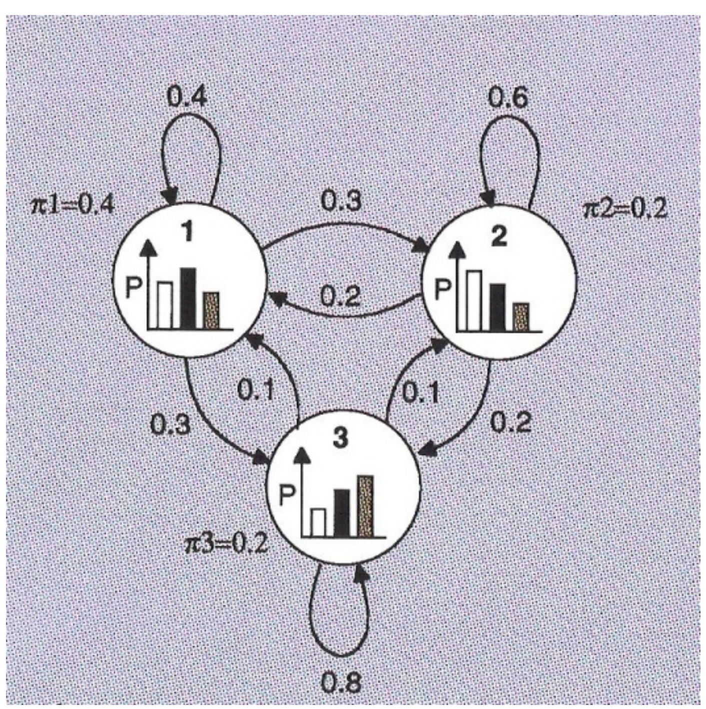

S
A
B
C
E


Emission distributions
q = A
q = B
q = C
p(x|q)
0.8
0.6
0.4
0.2
0
0
1
2
3
4
observation x


State sequence
AAAAAAAABBBBBBBBBBBCCCCBBBBBBBC
Observation sequence
x_n
0
1
2
3
10
20
30
time step n

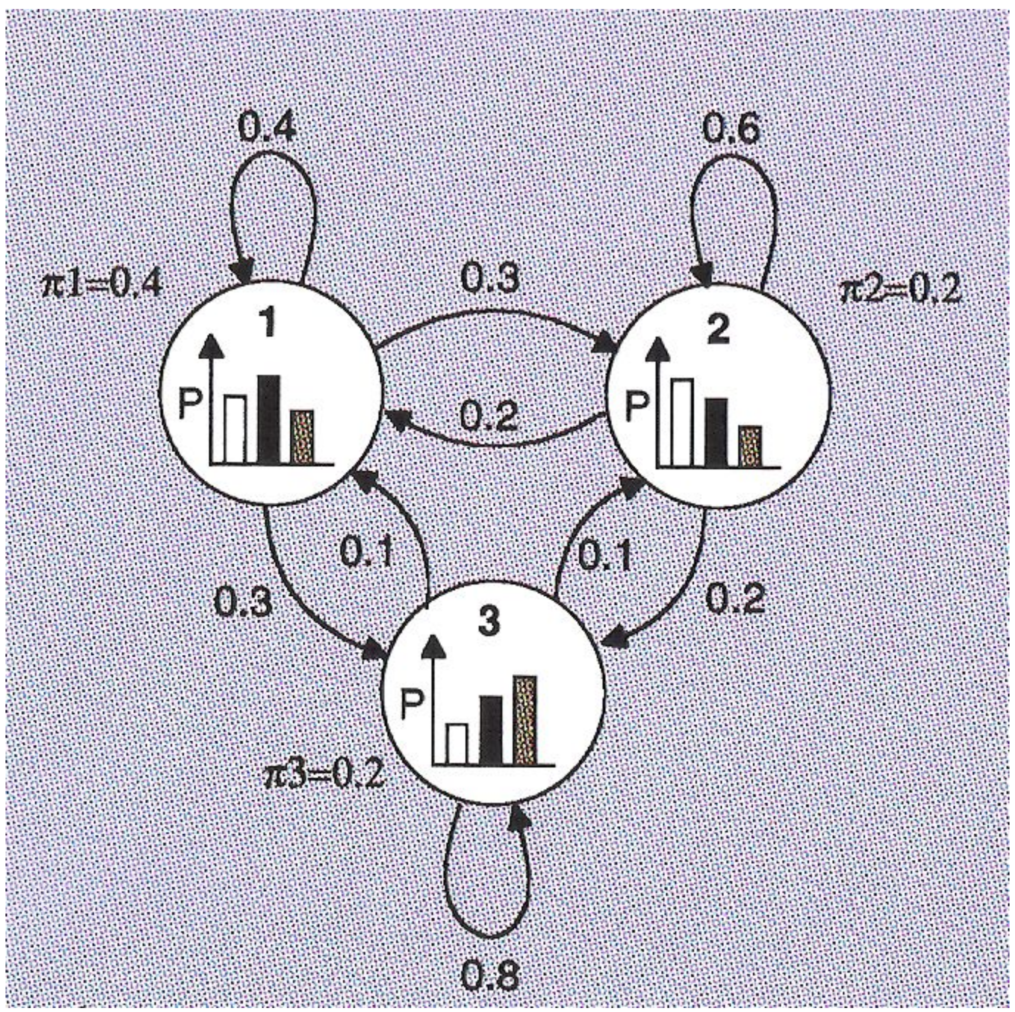
Probabilités d'émission





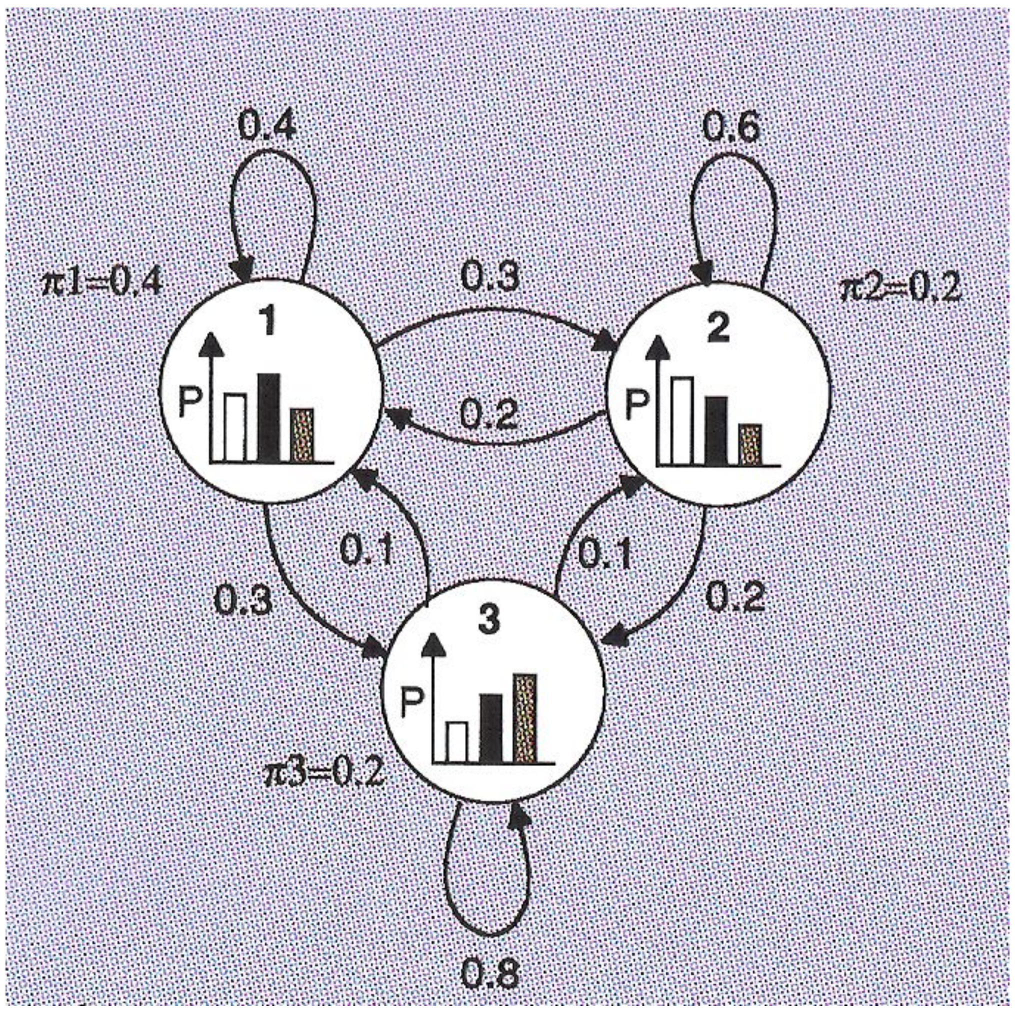
- Les observations sont continues
  - *p(x/q)* suit une distribution continue: gaussienne, multi-gaussienne, …
- Les observations sont discrètes:
  - *p(x/q)* est modélisé par des histogrammes

Probabilités d'émission



- *A*, l'ensemble des probabilités de transition
- *B*, l'ensemble des densités de probabilité d'observation
- π, l'ensemble des probabilités d'état initial
- $M=(A,B,\pi)$

Probabilités d'émission



q_t = état au temps t

$A = \{a_{ij}\}$ avec $a_{ij} = P(q_t = j \mid q_{t-1} = i)$

$B = \{b_j(x)\}$ avec $b_j(x) = p(x \mid q_t = j)$

$\pi = \{\pi_i\}$ avec $\pi_i = P(q_t = i)$

$M = (A, B, \pi)$

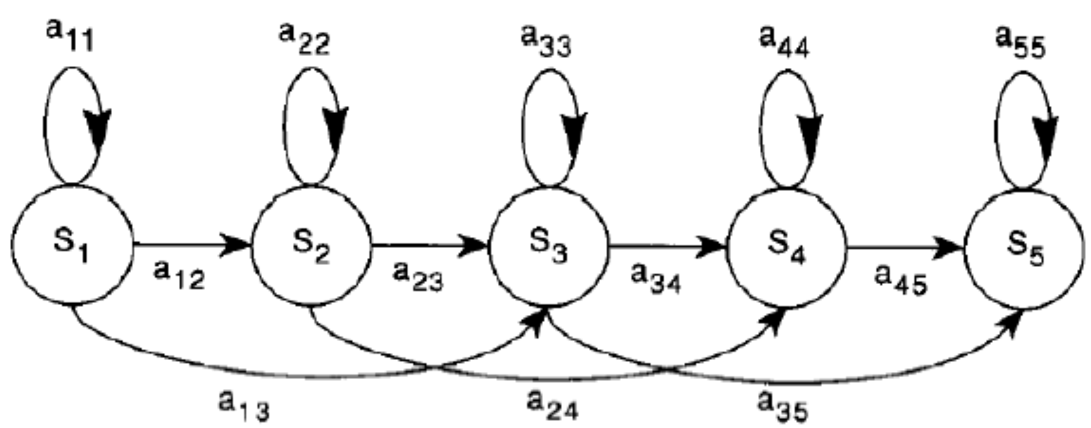
Séquence d'observation

$X = \{x_1, \ldots, x_n\}$

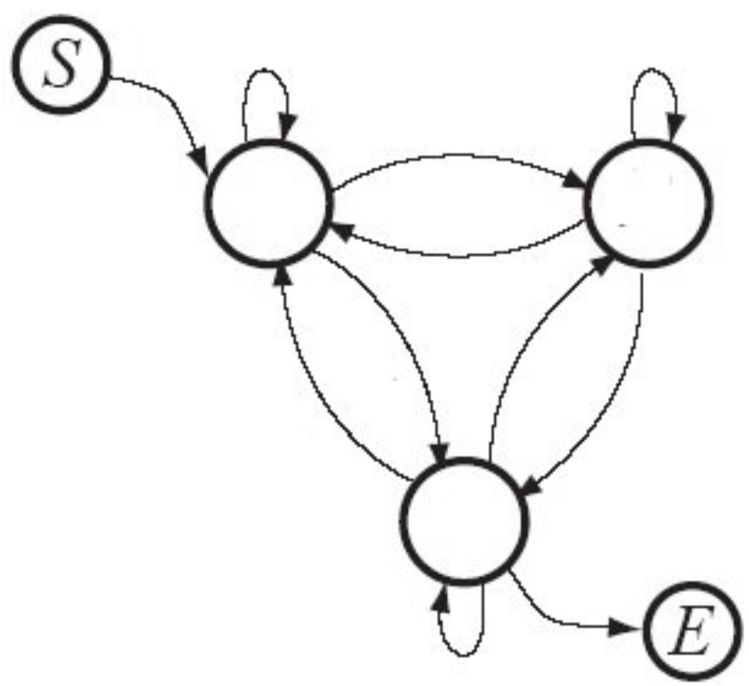
Séquence d'état particulière

$\mathrm{q} = \{q^1, \ldots, q^n\}$

Topologie des HMMs



Gauche-droite



Ergodique

LES PROBLÈMES QUE PEUVENT RÉSOUDRE LES HMMS

Problèmes

1. Génération d'une séquence d'observation X

2. Probabilité de la génération d'une séquence

- **connaissant** X **et** q
- **connaissant** X

3. Alignement: Séquence d'état optimale connaissant X

4. Réajustement des paramètres du HMM (A,B,π) **pour maximiser** $P(X|M)$

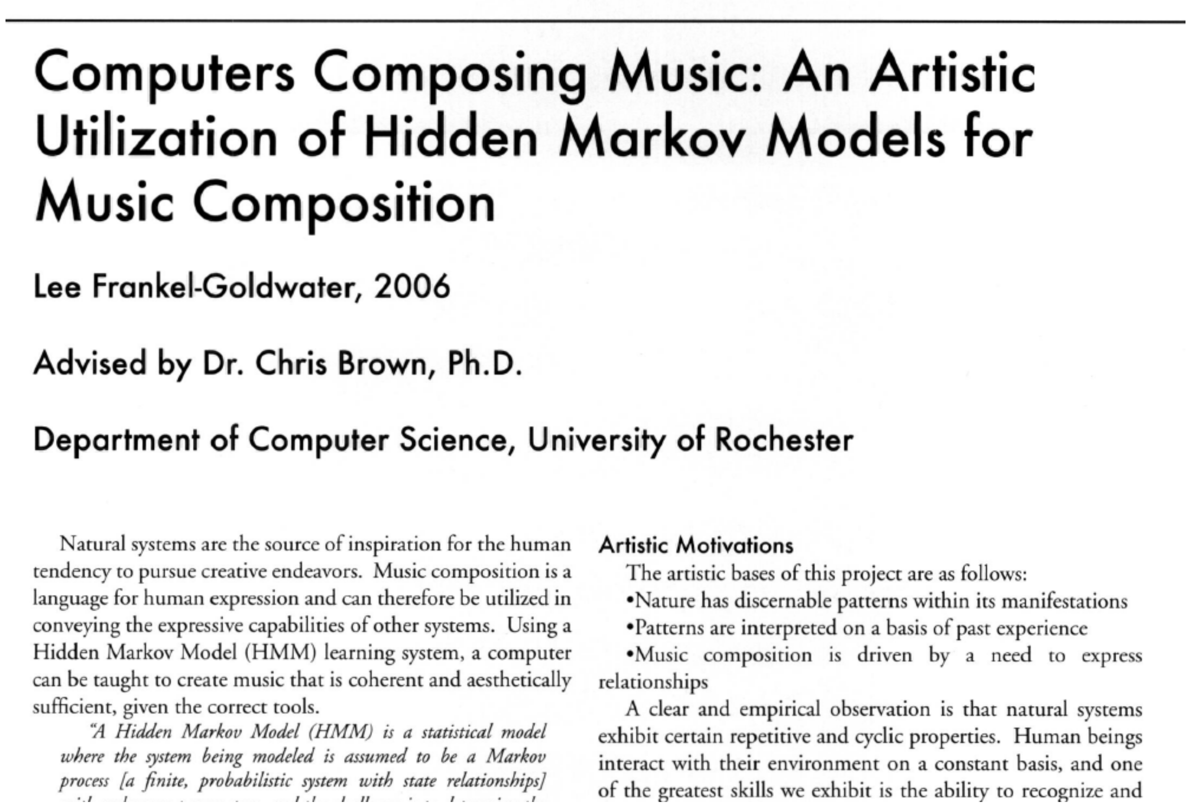
1. Génération d'une séquence d'observation X

- Le HMM peut être utilisé comme générateur de séquence
  - Les sauts d'états se font selon les probabilités de transition
  - Les observations sont émises selon leur loi de distribution dans l'état où l'on se trouve

- Exemple: distribution discrète dans des urnes

1. Génération d'une séquence d'observation X

Music Composition



Computers Composing Music: An Artistic Utilization of Hidden Markov Models for Music Composition

Lee Frankel-Goldwater, 2006

Advised by Dr. Chris Brown, Ph.D.

Department of Computer Science, University of Rochester

Natural systems are the source of inspiration for the human tendency to pursue creative endeavors. Music composition is a language for human expression and can therefore be utilized in conveying the expressive capabilities of other systems. Using a Hidden Markov Model (HMM) learning system, a computer can be taught to create music that is coherent and aesthetically sufficient, given the correct tools.

"A Hidden Markov Model (HMM) is a statistical model where the system being modeled is assumed to be a Markov process [a finite, probabilistic system with state relationships]

Artistic Motivations

The artistic bases of this project are as follows:

•Nature has discernable patterns within its manifestations

•Patterns are interpreted on a basis of past experience

•Music composition is driven by a need to express relationships

A clear and empirical observation is that natural systems exhibit certain repetitive and cyclic properties. Human beings interact with their environment on a constant basis, and one of the greatest skills we exhibit is the ability to recognize and

ex. http://youtu.be/UVuqYZDMlqA?t=1m22s

Text Creation

Ex. Mark V Shaney

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

"Mark V Shaney is a synthetic Usenet user whose postings in the net.singles newsgroups were generated by Markov chain techniques, based on text from other postings."

2.1. Probabilité de la génération d'une séquence connaissant X et q

- Hypothèses:
  - La séquence d'observation est connue
  - La séquence d'état est connue

$$X = \{x_1, x_2, \ldots, x_n\}$$
$$\mathrm{q} = \{q^1, q^2, \ldots, q^n\}$$

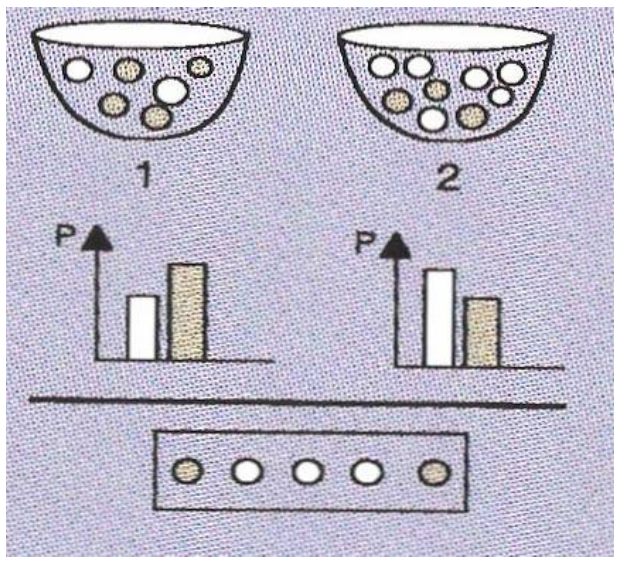
$$P(X, \mathrm{q} \mid M) = P(X \mid \mathrm{q}, M) P(\mathrm{q} \mid M)$$

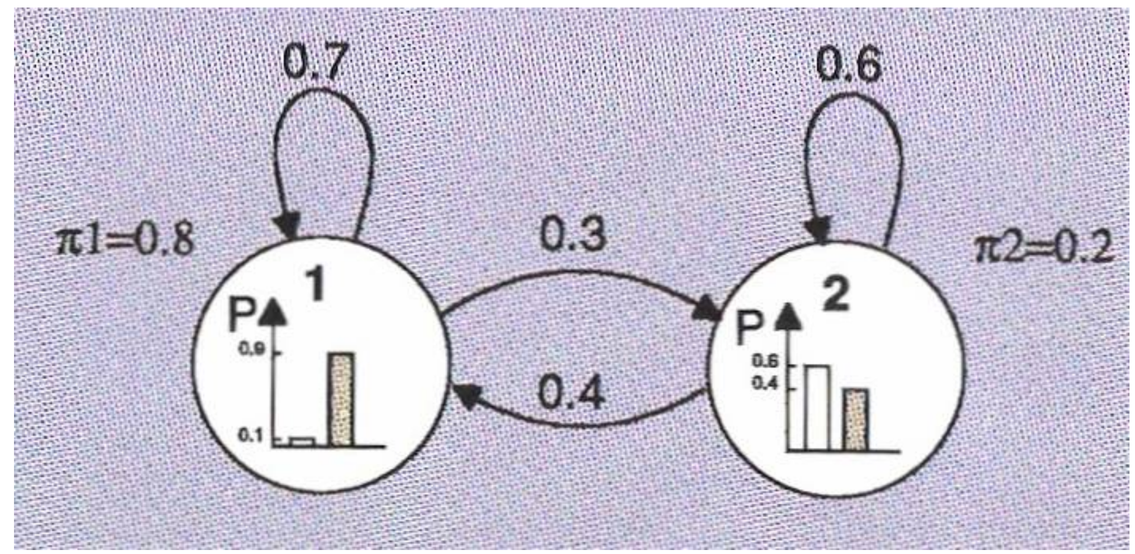
avec

$$P(X \mid \mathrm{q}, M) = b_{q^1}(x_1) b_{q2}(x_2) \ldots b_{q^n}(x_n)$$ ← Produit des probabilités d'émission

$$P(\mathrm{q} \mid M) = \pi_{q^1} a_{q^1 q^2} a_{q^2 q^3} \ldots a_{q^{n-1} q^n}$$ ← Produit des probabilités de transition

Exemple: probabilité de la génération d'une séquence par un HMM





$$\text{Suite d'observations } X = (g, b, b, b, g)$$

$$\text{Suite d'états } q = (1,1,2,2,1)$$

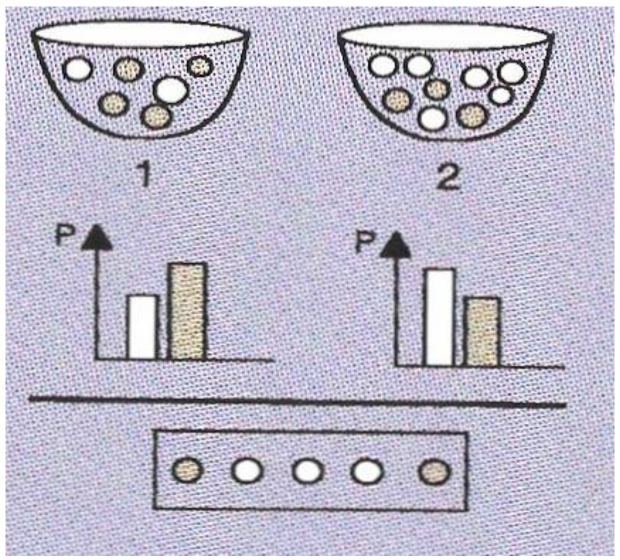
← **Hypothèses**

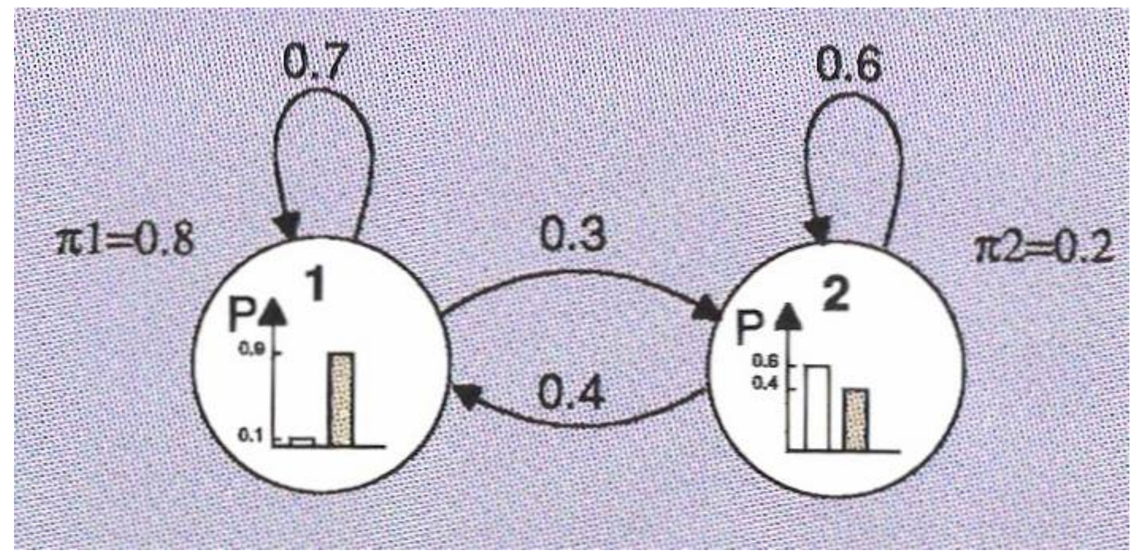
$$P(X, q \mid M) = \pi_1 b_1(g) a_{11} b_1(b) a_{12} b_2(b) a_{22} b_2(b) a_{21} b_1(g)$$

$$= 0.8 \cdot 0.9 \cdot 0.7 \cdot 0.1 \cdot 0.3 \cdot 0.6 \cdot 0.6 \cdot 0.6 \cdot 0.4 \cdot 0.9$$

$$= 0.0011757$$

Exemple: probabilité de la génération d'une séquence par un HMM





$$\text{Suite d'observations } \; X = (b, b, g, b, g)$$

$$\text{Suite d'états } \; q = (2,1,1,2,2)$$

← **Hypothèses**

$$P(X, q \mid M) = \pi_2 b_2(b) a_{21} b_1(b) a_{11} b_1(g) a_{12} b_2(b) a_{22} b_2(g)$$

$$= 0.2 \cdot 0.6 \cdot 0.4 \cdot 0.1 \cdot 0.7 \cdot 0.9 \cdot 0.3 \cdot 0.6 \cdot 0.6 \cdot 0.4$$

$$= 0.0001306368$$

2.2. Probabilité de la génération d'une séquence connaissant X

- Hypothèses:
  - La séquence d'observation est connue
  - La séquence d'état est inconnue

$$X = \{x_1, x_2, \ldots, x_n\}$$

$$P(X \mid M) = \sum_{\text{all q}} P(X \mid \text{q}, M) P(\text{q} \mid M)$$ ← **Somme sur tous les chemins possibles**

avec

$$P(X \mid \text{q}, M) = b_{q^1}(x_1) b_{q2}(x_2) \ldots b_{q^n}(x_n)$$

$$P(\text{q} \mid M) = \pi_{q^1} a_{q^1 q^2} a_{q^2 q^3} \ldots a_{q^{n-1} q^n}$$

Problème: grand nombre de chemins. Complexité proportionnelle à $2nN^n$. Exemple: 10^{72} calculs pour N=5, n=100, HMM ergodique

2.2. Probabilité de la génération d'une séquence connaissant X

- Il existe des récursions pour calculer de façon plus optimale cette probabilité
  - Procédure **forward**
  - Procédure **backward**
- Complexité proportionnelle à $N^2 n$.
- Exemple: 3000 calculs pour N=5, n=100, HMM ergodique

On considère la variable "forward"

$$\alpha_{\mathrm{n}}(i) = P(x_1 x_2 \ldots x_n, q_n = i \mid M)$$

1. Initialisation

$$\alpha_1(i) = \pi_i b_i(x_1)$$

2. Induction

$$\alpha_{\mathrm{t}+1}(j) = \left[\sum_{i=1}^{Q} \alpha_{\mathrm{t}}(i) a_{ij}\right] b_j(x_{t+1})$$

3.Terminaison

$$\mathrm{P}(\mathrm{X} \mid \mathrm{M}) = \sum_{i=1}^{Q} \alpha_{\mathrm{N}}(i)$$

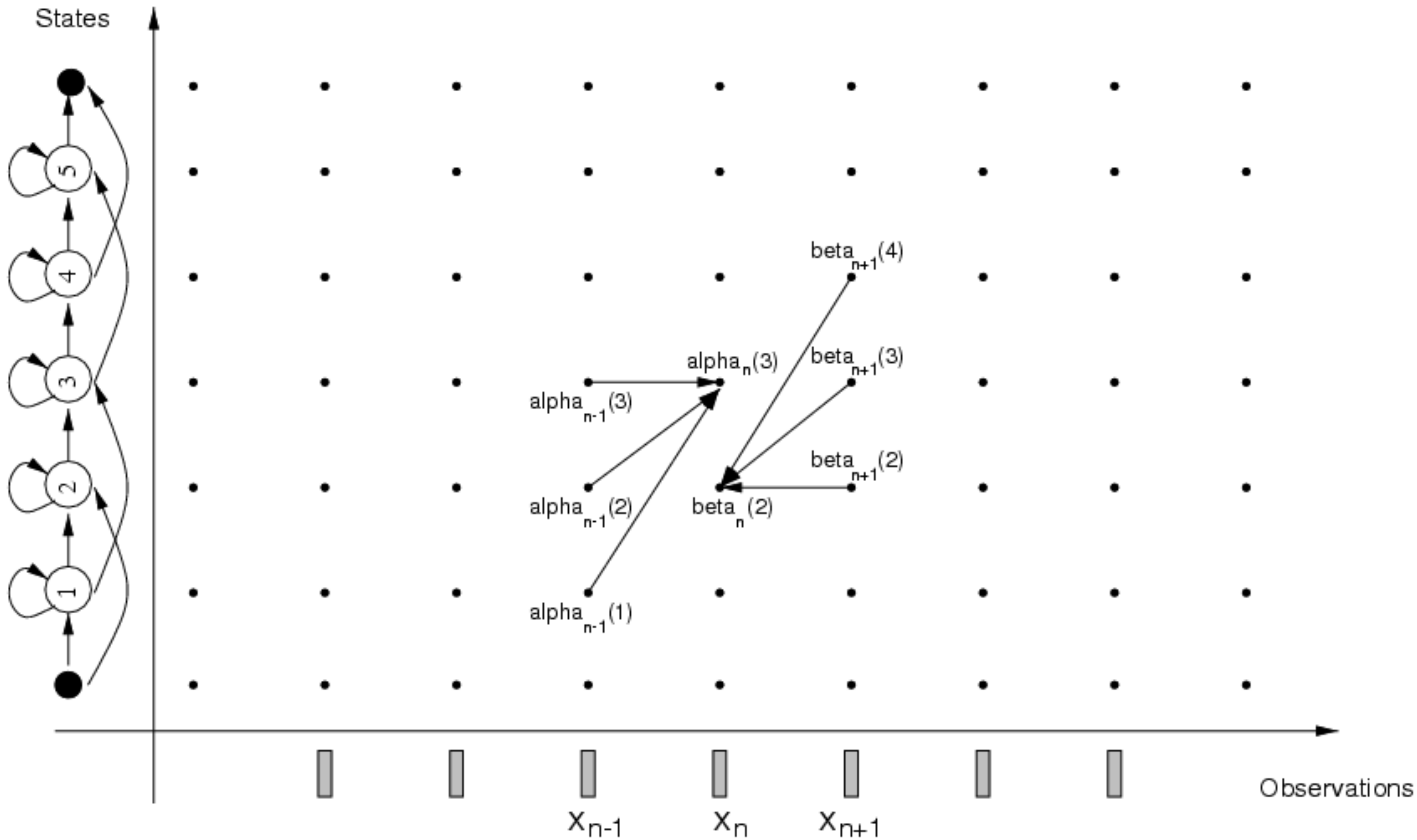


Figure 2.8: Forward and backward recursion. The HMM topology is left-right but permits in this case to skip the next state and to go to the second next state. The forward recursion propagates from left to right to compute the α probabilities. For example, the lattice point corresponding to $\alpha_n(3)$ can only be reached from states q_1, q_2 and q_3, and the recursion in this point is $\alpha_n(3) = [\alpha_{n-1}(1)p(q_3|q_1) + \alpha_{n-1}(2)p(q_3|q_2) + \alpha_{n-1}(3)p(q_3|q_3)]\, p(x_n|q_3)$. The backward probabilities are computed in a similar manner but propagating the recursion from right to left. Any HMM topology can be taken into account, the beam of forward/backward connections getting larger with more intricated topologies.

3. Alignement: séquence d'état optimale connaissant X

- Hypothèse
  - La séquence d'observations est connue
  - La séquence d'états est inconnue
- On veut déterminer la séquence d'états qui maximise la probabilité que le modèle génère l'observation: $P(X,Q_{max}|M)$
- Algorithme de Viterbi:

On considère la variable "forward"

$$\delta_{\mathrm{n}}(i) = \max_{q_1, q_2, \ldots, q_{n-1}} P(x_1 x_2 \ldots x_n, q_1 q_2 \ldots q_{n-1}, q_n = i \mid M)$$

1. Initialisation

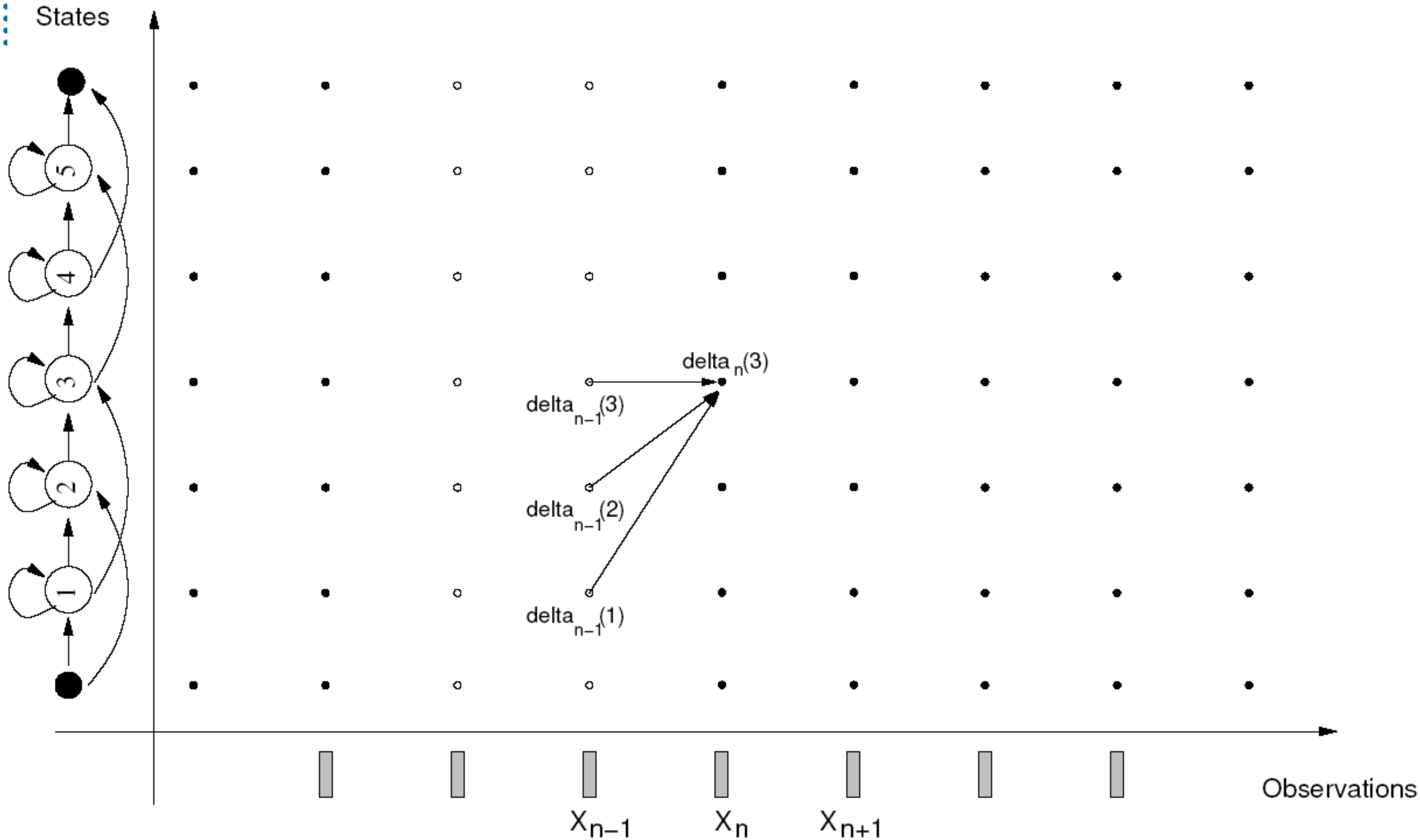
$$\delta_1(i) = \pi_i b_i(x_1)$$

2. Récursion

$$\delta_{\mathrm{t}}(j) = \max_{1 \le i \le Q} \left[\delta_{t-1}(i) a_{ij} \right] b_j(x_t)$$

3. Terminaison

$$\mathrm{P}(\mathrm{X},\mathrm{Q}_{\max} | \mathrm{M}) = \max_{1 \le i \le Q} [\delta_{\mathrm{N}}(i)]$$



Viterbi algorithm propagates from left to right to compute the deltas. In this figure, $delta_n(3)$ is computed by looking at the highest value of the potential preceding deltas times the corresponding transition probabilities a_{ij}. In the example illustrated above, $delta_{n-1}(2).a_{23} > delta_{n-1}(3).a_{33}$ and $delta_{n-1}(2).a_{23} > delta_{n-1}(1).a_{13}$. The best transitions leading from one point of the lattice to the next are then kept in memory to allow backtracking from the last state to the first.

4. Réajustement des paramètres du HMM (A, B, π) pour maximiser $\mathrm{P}(X \mid M)$

Pas de solution optimale !

Méthode utilisant le critère Viterbi:

1. On part d'un HMM avec un set de paramètres initiaux
2. On aligne l'ensemble des observations avec Viterbi
3. On accumule les ensembles d'observations dans les états
4. On ré-estime
   - Les probabilités d'émission
   - Les probabilités de transition
5. Retour en 2 et on itère jusqu'à la convergence

Plus de détails la semaine prochaine!

HMM, outil universel?

- Non, mais s'applique à beaucoup de problèmes!
- S'applique lorsque le système à modéliser génère des **séquences** de données et lorsqu'il y a plusieurs **modes de génération** de ces séquences qui correspondent à des **états** différents du système.
- Exemples: signal de parole, écriture cursive, séquence ADN, prédiction financière

What you should know

- Intervalles de confiance
  - Motivation & définition
- Modèles de Markov Cachés
  - Modèles de Markov Vs Modèles de Markov cachés
- Les paramètres des HMM
  - A, B, π, M
- Les problèmes que peuvent résoudre les HMMs
  - Génération d’une séquence d’observation X
  - Probabilité de la génération d’une séquence connaissant X et q
  - Probabilité de la génération d’une séquence connaissant X
  - Alignement: séquence d’état optimale connaissant X

References

[1]L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[2] HTKBook - http://htk.eng.cam.ac.uk/docs/docs.shtml