HIDDEN *semi*-MARKOV MODEL: forget (half of) Markov hypothesis ?!

LILIAN BESSON & VALENTIN BRUNCK

PGM course for the MVA master at ENS CACHAN *(2015-16)*

INTRODUCTION

- Weakness of HMM: **Geometric durations**

$$\mathbb{P}(d|j) = (1 - A_{j,j}) A_{j,j}^d.$$

- HSMM: models the time spent on a hidden state (sequence duration).
- **Practical application:** music sheet matching & alignment.

THE MODEL

HSMM $\equiv$ **HMM without the Markov property**.
Hidden state q_t:

- **HMM:** $q_t \to u_t$
- **HSMM:** $q_t \to (u_{1:d_t})_t$ (sequence)

$$\mathbb{P}\Big((q_n, (u_{1:d_n})_n), n = 1..N\Big) = \mathbb{P}(q_1)\mathbb{P}((u_{1:d_1})_1|q_1) \cdots \prod_{n=2..N} \Big(\mathbb{P}(q_n|q_{n-1})\mathbb{P}((u_{1:d_n})_n|q_n)\Big)$$

FILTERING (α, β)

Forward $\alpha-$recursion:

$$\alpha_{t,j} \equiv \sum_{d=1}^{d_{\max}^{(j)}} B_{t,j,d} \mathbf{D_{j,d}} \left(\sum_{i=1}^{K} A_{i,j} \alpha_{t-d,i} \right)$$

backward $\beta-$recursion:

$$\beta_{t,i} \equiv \sum_{j=1}^{K} A_{i,j} \left(\sum_{d=1}^{d_{\max}^{(j)}} \mathbf{D_{j,d}} B_{t+d,j,d} \beta_{t+d,j} \right)$$

In practice: logs (and `logsumexp`) to avoid underflow errors. Complexity: $O(T D_{\max} K^2)$.

ILLUSTRATION

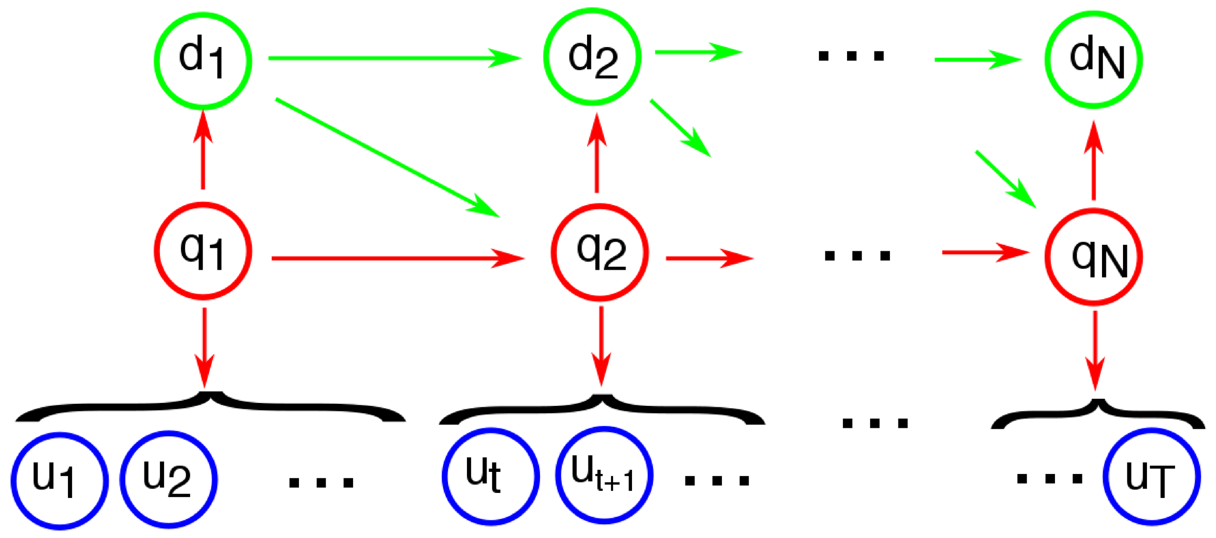

Figure 1: N sequence of observations of a HSMM (hidden state: $q_k \in \{1..K\}$, observed variables: u_t, duration of i^{th} sequence: d_i).

ASSUMPTIONS

$$\begin{cases} A_{i,j} = \mathbb{P}(j|i) & \text{(Markovian transition } i \rightarrowtail j) \\ D_{j,d} = \mathbb{P}(d|j) & (\equiv (1 - A_{j,j}) A_{j,j}^d \text{ for HMM}) \\ B_{t,j,d} = \mathbb{P}(u_{t-d+1:t}|j,d) = \prod_{t'=t-d+1}^{t} \mathcal{N}(y_{t'}|j,d) \end{cases}$$

INFERENCE (E-M)

$$\begin{cases} \widehat{\pi}_i^{\text{new}} \leftarrow & \mathbb{P}(q_0 = i | u_0, \dots, u_T) \\ \widehat{A}_{i,j}^{\text{new}} \leftarrow & \sum_t \xi_t(i,j) / \left(\sum_{i \neq j} \sum_t \xi_t(i,j) \right) \\ \widehat{D}_{j,d}^{\text{new}} \leftarrow & \eta(j,d) / \left(\sum_d \eta(j,d) \right) \end{cases}$$

$$\xi_t(i,j) = \mathbb{P}(q_t = i, q_{t+1} = j | u_0..u_T)$$

$$\eta(j,d) = \mathbb{P}(q_{d-v} = j, v = 1..d, q_d \neq j | u_1..u_T) + \sum_{t=1..T} \mathbb{P}(q_t \neq j, q_{d-v} = j, v = 1..d, q_{t+u+1} \neq j | u_1..u_T)$$

This methods complexity is: $m \times O(T K^2 D_{\max}^2)$ (for m steps).

CONCLUSION *Thanks for reading!*

- HMM are a special case of HSMM, and our HSMM implementation can emulate a HMM,
- For both HMM and HSMM, $\alpha - \beta$ is tractable and efficient **for truncated** $D_{\max}$,
- For Geometric durations, E-M for HSMM is *very* similar to E-M for HMM (cf. HMK3),
- But for other durations distribution, E-M is more complicated, but works in practice (cf. plots).

GEOMETRIC

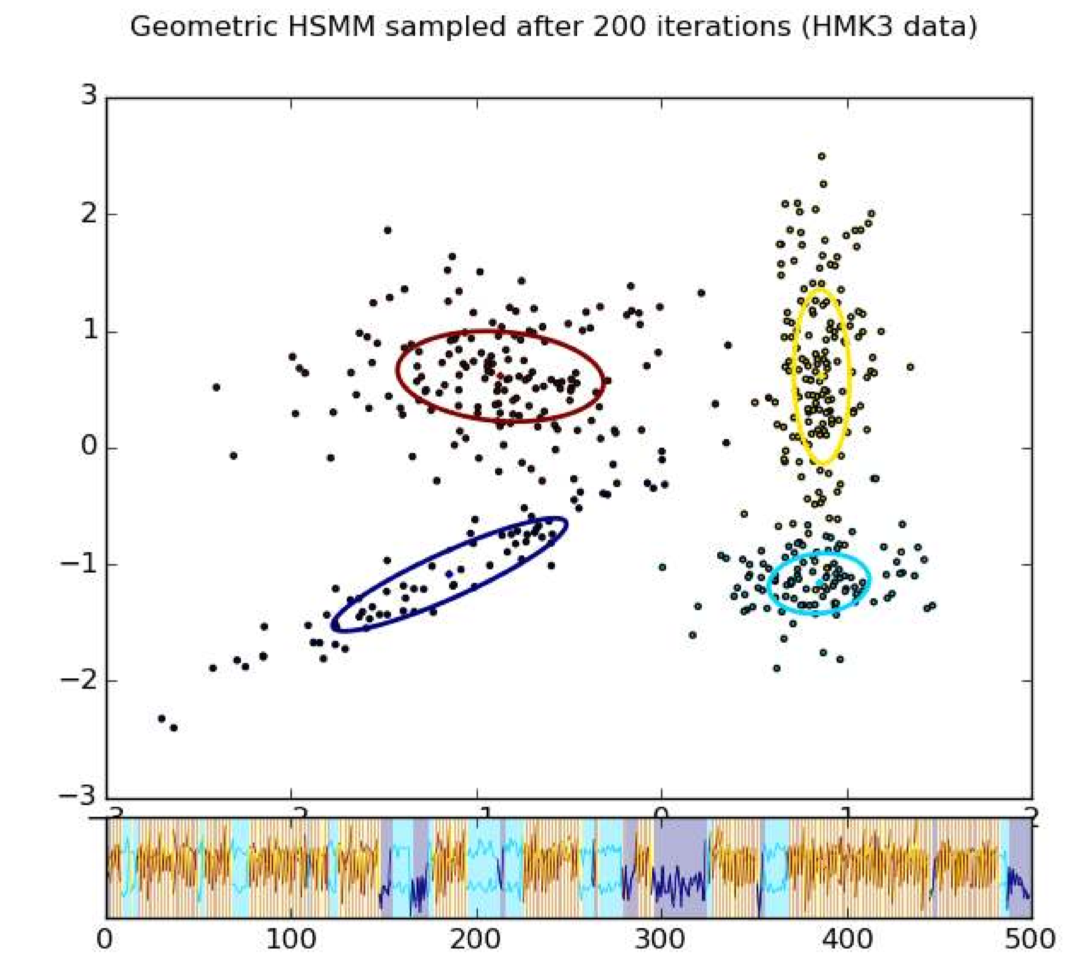

Figure 2: HSMM with Geometric d distribution

POISSON

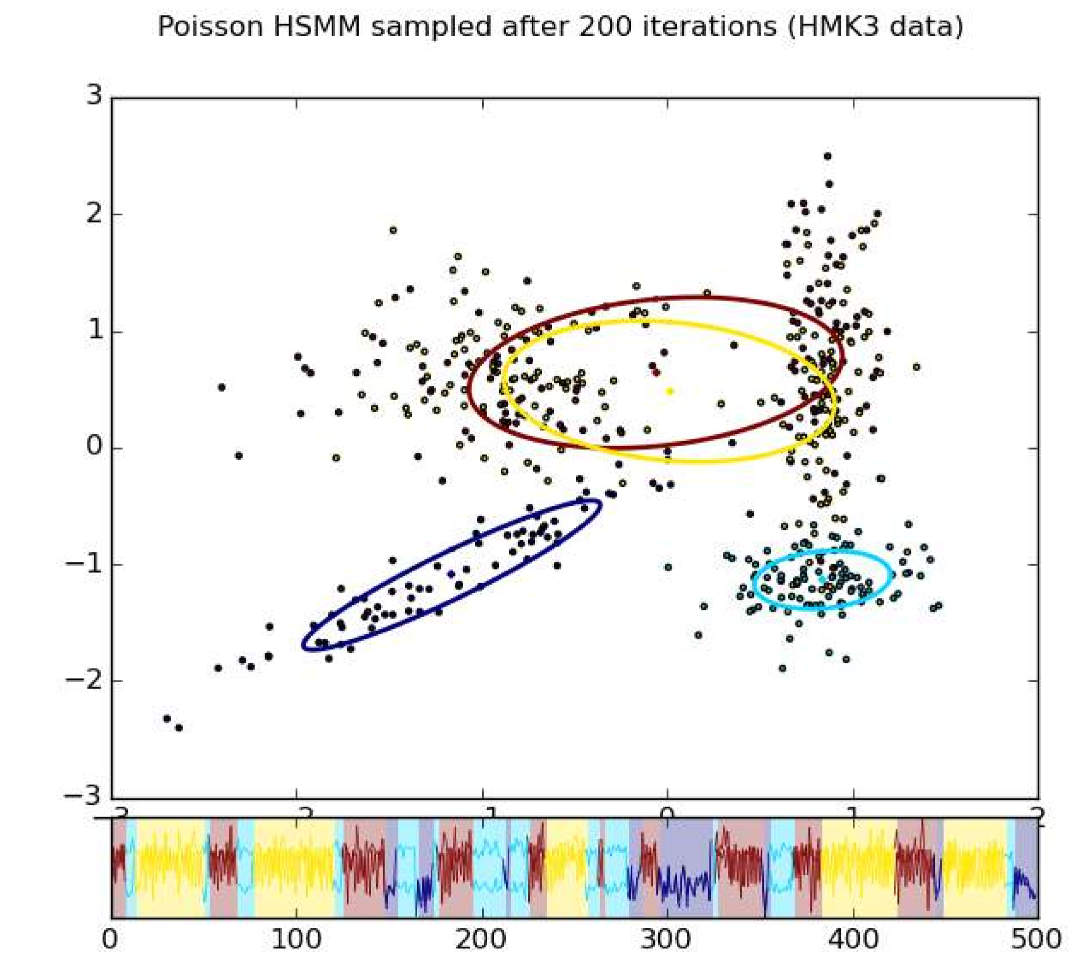

Figure 3: HSMM with Poisson d distribution

POISSON (WORKS!)

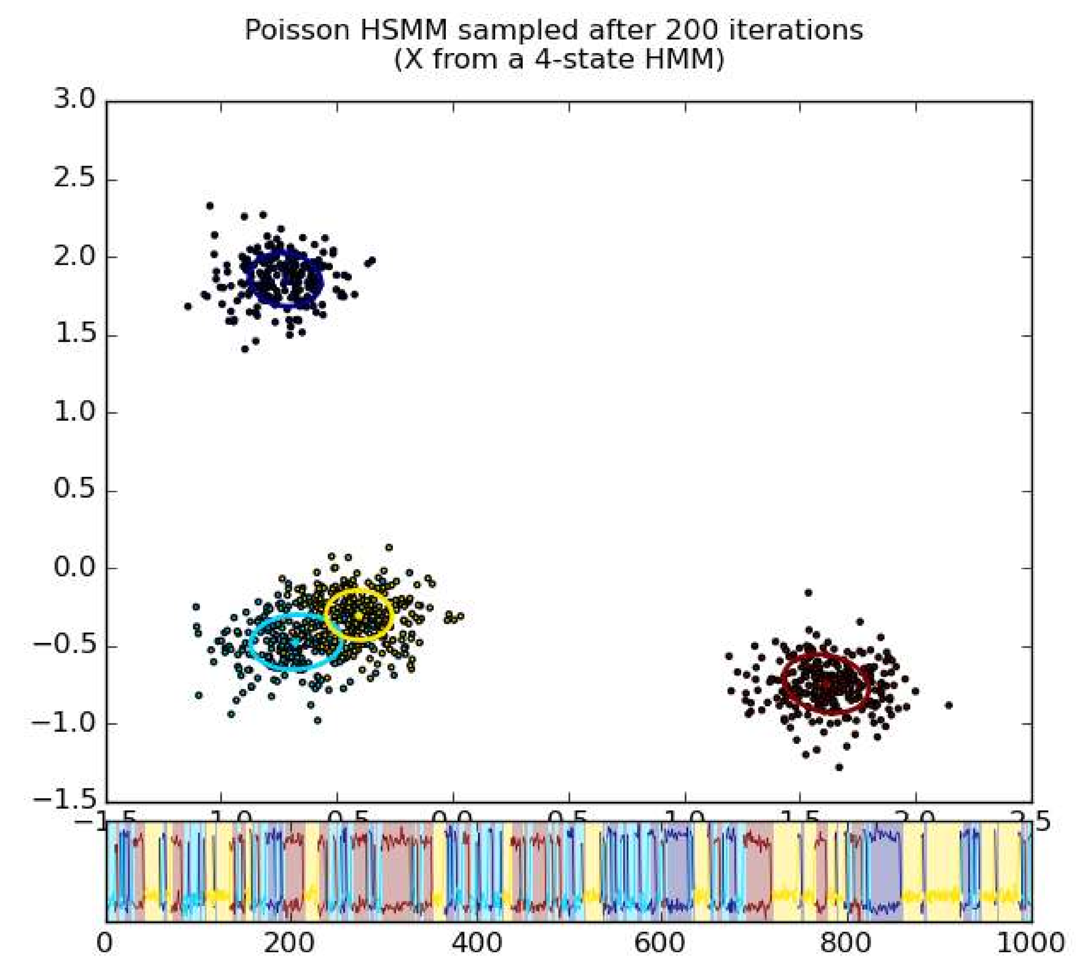

Figure 4: $2D$ data drawn from a 4-state HMM

Note: **No** transition between state/cluster 1 and 3, but almost the same Gaussian (μ_i, Σ_i).

POISSON (FAILS!)

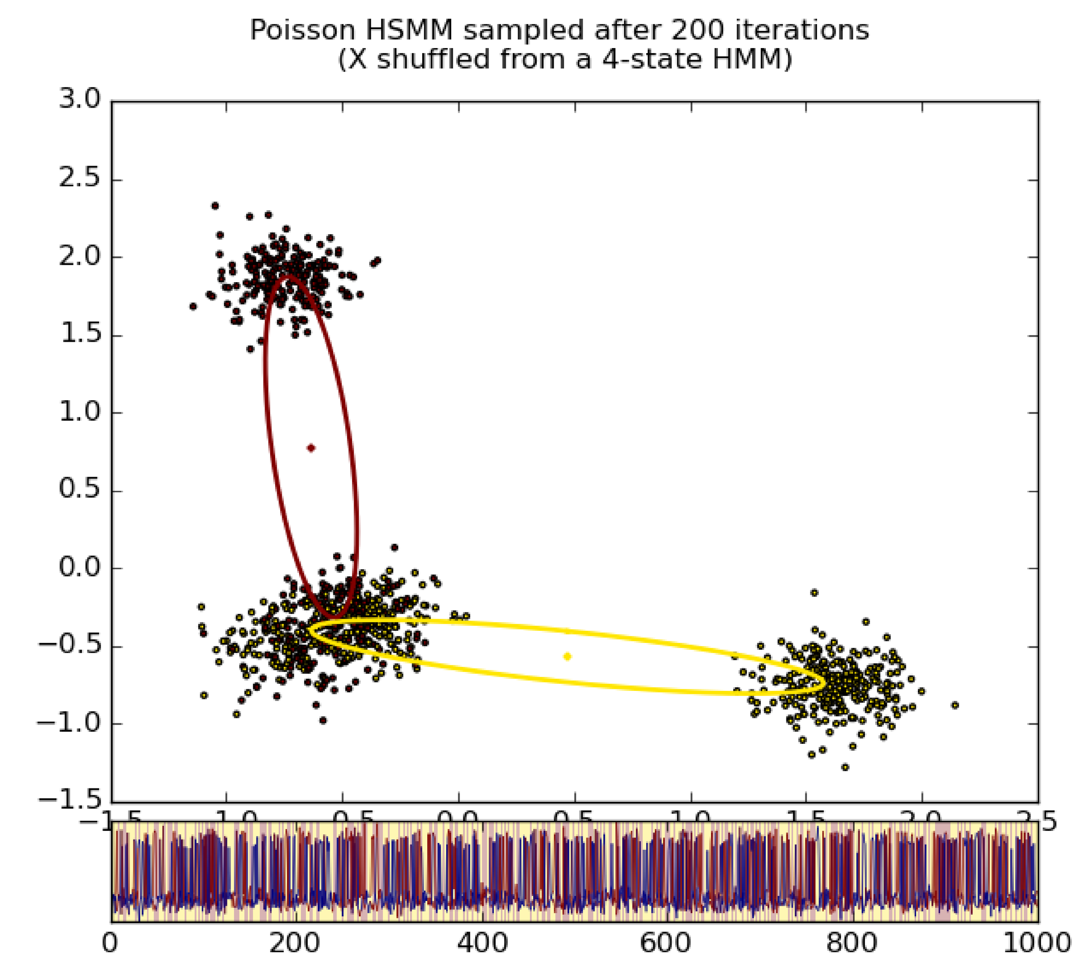

Figure 5: Same data, but shuffled