# 11. Hidden Markov Models on acyclic graphs

## A. The model

**Definition 1a** Let $T = (V,E)$ be an undirected acyclic graph and $s_i$, $i \in V$ be $K$-valued discrete random variables indexed by its vertices. A p.d. for configurations $s \in K^{|V|}$ is a Markov model if the following holds. By fixing an arbitrary vertex $i_* \in V$ and denoting by $c(i)$ the next vertex in the path connecting $i \in V$ with $i_*$, the p.d. can be written as

$$p(s) = p(s_{i_*}) \prod_{i \neq i_*} p(s_i \mid s_{c(i)}) \qquad \blacksquare$$

**Definition 1b** Equivalently, a p.d. $p(s)$ for $s \in K^{|V|}$ is a Markov model if

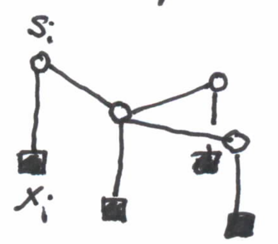
$$p(s) = \prod_{ij \in E} p(s_i, s_j) \Big/ \prod_{i \in V} p^{n_i - 1}(s_i)$$

holds, where $n_i$ is the degree of vertex $i \in V$ $\blacksquare$

An HMM on an (undirected) acyclic graph $T = (V,E)$ is a p.d. for pairs $s \in K^{|V|}$, $x \in F^{|V|}$, where $F$ is a feature space and $p(x,s) = p(x \mid s)\, p(s)$ where

$$p(x \mid s) = \prod_{i \in V} p(x_i \mid s_i) \quad \text{(conditional independence)}$$

$p(s)$ - Markov model on $T$



## B. Inference tasks

Consider the following task: Given an observation field $x \in F^{|V|}$, compute

$$p(x) = \sum_{s \in K^{|V|}}' p(x,s)$$

By substituting the model (Def. 1a) and denoting the chosen vertex by $r \in V$ (root) we get

$$p(x) = \sum_{s \in K^{|V|}}' p(s_r) p(x_r | s_r) \prod_{j \in V \setminus r} p(s_j | s_{c(j)}) p(x_j | s_j)$$

The observations $x$ are fixed $\Rightarrow$ the problem has the form

$$\sum_{s \in K^{|V|}}' \prod_{i \in V} \varphi_i(s_i) \prod_{j \in V \setminus r} f_j(s_j, s_{c(j)})$$

where

$$\varphi_r(s_r) = p(s_r) p(x_r | s_r)$$

$$\varphi_i(s_i) \equiv 1 \quad \forall\, i \neq r$$

$$f_j(s_j, s_{c(j)}) = p(s_j | s_{c(j)}) \cdot p(x_j | s_j)$$

The algorithm recomputes the $\varphi$-s starting from an arbitrary leaf $b \in V$

$$\varphi_{c(b)}(s_{c(b)}) := \varphi_{c(b)}(s_{c(b)}) \cdot \sum_{s_b \in K}' f_b(s_b, s_{c(b)}) \cdot \varphi_b(s_b)$$

and „removes" $b$. This is repeated until only the root remains. Finally

$$p(x) = \sum_{k \in K}' \varphi_r(k)$$

Complexity of the algorithm: $|K|^2 |E|$

**Remark 1**

The same approach can be applied for the task

$$s_* = \arg\max_{s \in K^{|V|}} \log p(x, s)$$

simply by replacing operations

$$\times \rightarrow +$$
$$+ \rightarrow \max$$

**1C. Computing marginals**

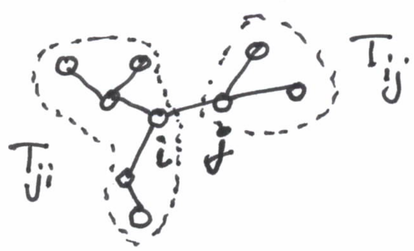
Consider the following task: Given ~~the~~ an observation field $x \in F^{|V|}$, compute marginal prob's

$$p(x, s_i), \quad \forall\, i \in V, \quad s_i \in K$$

Remember the algorithm for computing marginals for an HMM on a chain discussed in sec. 5

**Here:**

- Define a subtree $T_{ij}$ for each oriented edge $ij \in E$. It is composed of all vertices for which the path connecting the vertex with vertex $i$ goes through vertex $j$



- It follows from Def. 16 that $p(x, s_i)$ can be expressed as

$$p(x, s_i) = p(s_i)\, p(x_i \mid s_i) \prod_{j \in \mathcal{N}_i} \underbrace{p(x_{T_{ij}} \mid s_i)}_{\varphi_{ij}(s_i)}$$

- The $\varphi$'s fulfil the following system of equations

$$\varphi_{ij}(s_i) = \sum_{s_j} p(s_j \mid s_i)\, p(x_j \mid s_j) \prod_{\substack{l \in \mathcal{N}_j \\ l \neq i}} \varphi_{jl}(s_j)$$

- Two passes through all edges of $T$ are needed in order to compute all of them $\Rightarrow$ complexity $2|K|^2|E|$

## D. Learning the tree structure

Given: i.i.d. sample of state configurations $s \in K^{|V|}$ generated by a Markov model on a tree

Question: Can we estimate its parameters and its structure?

Denote by $\beta(s)$ the empirical prob. assoc. with the sample. The log-likelihood for known structure $(V, E)$

$$\sum_{s\in K^{|V|}}' \beta(s)\log p(s) \overset{\text{Def. 16}}{=} \dots\dots$$

$$= \sum_{ij\in E}' \sum_{s_i,s_j}' \beta(s_i,s_j)\log p(s_i,s_j) - \sum_{i\in V}' (n_i-1)\sum_{s_i}' \beta(s_i)\log p(s_i)$$

$\to \max\limits_{p}$

Maximiser given by $p(s_i)=\beta(s_i)$, $p(s_i,s_j)=\beta(s_i,s_j)$

It remains to solve

$$\sum_{ij\in E}' \sum_{s_i,s_j}' \beta(s_i,s_j)\log \beta(s_i,s_j) - \sum_{i\in V}' (n_i-1)\sum_{s_i}' \beta(s_i)\log \beta(s_i) \to \max_{E:\,\text{tree}}$$

Both terms depend on the unknown structure. Rewrite the objective

$$\sum_{ij\in E}' \Big\{ \sum_{s_i,s_j}' \beta(s_i,s_j)\log\beta(s_i,s_j) - \sum_{s_i}' \beta(s_i)\log\beta(s_i) - \sum_{s_j}' \beta(s_j)\log\beta(s_j) \Big\} +$$

$$+ \sum_{i\in V}' \sum_{s_i}' \beta(s_i)\log\beta(s_i)$$

The last term does not depend on the edge structure $E$. Denoting the values of the curly brackets in the first term by $h_{ij}$, we get

$$\sum_{ij\in E}' h_{ij} \to \max_{E:\,\text{tree}}$$

$\Rightarrow$ Maximum spanning tree problem; can be easily solved.