Ch. I Markov Models on chains and acyclic graphs

1. Markov models on chains

1A. Definitions and basic properties

- Sequence $S = (s_1, \ldots, s_n)$ of K-valued random variables $s_i \in K$
- K is a finite set, its elements are called states
- $p(s) = p(s_1, \ldots, s_n)$ - joint probability distribution on K^n

W.l.o.g. we may write

$$
\begin{aligned}
p(s_1, \ldots, s_n) &= p(s_n \mid s_1, \ldots, s_{n-1})\, p(s_1, \ldots, s_{n-1}) \\
&= \ldots \\
&= p(s_n \mid s_1, \ldots, s_{n-1})\, p(s_{n-1} \mid s_1, \ldots, s_{n-2}) \cdot \ldots \cdot p(s_2 \mid s_1)\, p(s_1)
\end{aligned}
$$

Definition 1a A p.d. on K^n is a Markov chain if

$$p(s) = p(s_1) \prod_{i=2}^{n} p(s_i \mid s_{i-1})$$

holds for $\forall\, s \in K^n$.

Definition 1b A p.d. on K^n is a Markov chain if

$$p(s) = \prod_{i=2}^{n} g_i(s_{i-1}, s_i)$$

holds for $\forall\, s \in K^n$, where $g_i : K^2 \to \mathbb{R}_+$ are some functions.

Equivalence:

a) $\to$ b) trivial

b) → a) recursively apply

$$p(s_{n-1}, s_n) = \left\{ \sum_{s_1,..,s_{n-2}} \prod_{i=2}^{n-1} g_i(s_{i-1}, s_i) \right\} g_n(s_{n-1}, s_n)$$

$\hookrightarrow g_n(s_{n-1}, s_n) = p(s_n | s_{n-1}) \cdot b_{n-1}(s_{n-1})$ with some b_{n-1}

Therefore

$$p(s_1,..,s_n) = \underbrace{\left[\prod_{i=2}^{n-1} g_i(s_{i-1}, s_i) \right] b_{n-1}(s_{n-1})}_{p(s_1,..,s_{n-1})} \cdot p(s_n | s_{n-1})$$

Another useful formula

$$p(s_1,.., s_n) = \frac{p(s_1, s_2) \cdot p(s_2, s_3) \cdot \ldots \cdot p(s_{n-1}, s_n)}{p(s_2) \cdot p(s_3) \cdot \ldots \cdot p(s_{n-1})}$$

Example 1 (Ehrenfest model)

The model considers N particles in two containers. At each discrete time $t = 1, 2, ...$, independently of the past, a particle is selected at random and moved to the other container. Let s_t denote the number of particles in the first container at time t. Then we have

$$p(s_t = k \mid s_{t-1} = \ell) = \begin{cases} \frac{N-\ell}{N} & \text{if } k = \ell + 1 \\ \frac{\ell}{N} & \text{if } k = \ell - 1 \\ 0 & \text{otherwise} \end{cases}$$

Q: How does $p(s_t = k)$, $k = 0, 1, .., N$ behaves for $t \to \infty$?

Example 2 (random walk on a graph)

Consider a random walk on an undirected graph V, E

- $K = V$ states, $s_t \in V$ - position of the walker at time t.
- $p(s_1)$ - some p.d. for the start vertex

- $$p(s_t = i \mid s_{t-1} = j) = \begin{cases} W_{ij} & \text{if } \{i,j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

where $W_{ij} \geq 0$ fulfill $\sum_{i \in N(j)} W_{ij} = 1 \quad \forall j \in V$ ∎

3. Homogeneous Markov chains, stationary p.d.s

A Markov chain is homogeneous if the conditional prob's $p(s_i \mid s_{i-1})$ do not depend on the position i, i.e.

$$p(s_i = k \mid s_{i-1} = k') = g(k, k') \quad \forall\, i = 2, \dots, n.$$

We know that

$$p(s_i = k) = \sum_{k' \in K} p(s_i = k \mid s_{i-1} = k')\, p(s_{i-1} = k').$$

Consider $p(s_i = k)$, $k \in K$ as components of an vector $\vec{\pi}_i \in \mathbb{R}^K_+$ and $p(s_i = k \mid s_{i-1} = k')$, $k, k' \in K$ as elements of a $K \times K$ matrix P. The previous eq. reads

$$\vec{\pi}_i = P \cdot \vec{\pi}_{i-1},$$

and, more general, we have $\vec{\pi}_i = P^{i-1} \vec{\pi}_1$. It may happen that there exists a p.d. $\vec{\pi}^*$ on K s.t. $P \cdot \vec{\pi}^* = \vec{\pi}^*$. We call it a stationary p.d.

Definition 2 A homogeneous Markov chain is irreducible if for each pair $k, k' \in K$ there is an $m > 0$ s.t. $P^m_{kk'} > 0$. I.e., there is a non-zero probability to reach state k starting from state k' (after m transitions). ∎

A condition somewhat stronger than irreducibility ensures existence and uniqueness of a stationary p.d.

Theorem 1 (w/o proof) If for some $m>0$ all elements of the matrix P^m are strictly positive, then the Markov chain has a unique stationary distribution $\vec{\pi}^*$, which is a fixpoint

$$P^n \cdot \vec{\pi} \xrightarrow{n\to\infty} \vec{\pi}^* \quad \forall \vec{\pi}.$$

Moreover

$$P^n = \vec{\pi}^* \otimes \vec{e} + \mathcal{E}(n),$$

where $\vec{e} = (1,\dots,1)$ and $\mathcal{E}_{kk'}(n) = \mathcal{O}(h^n)$ with some $0<h<1$. ∎

Remark 1 (infinite Markov chains)

- Consider infinite sequences $s = (s_1, s_2, \dots)$, $s_i \in K$. $K^{\mathbb{N}}$ is uncountable infinite. Any probability on $K^{\mathbb{N}}$ will assign zero probability to almost every sequence $s \in K^{\mathbb{N}}$
- A finite sequence $(k_1, k_2, \dots, k_n) \in K^n$ can be seen as a set of infinite sequences

$$(k_1,\dots,k_n) \longmapsto \{ s \in K^{\mathbb{N}} \mid s_1 = k_1, \dots, s_n \in k_n \}.$$

A Markov model on $K^{\mathbb{N}}$ assigns probabilities to such sets in the same way as described for finite sequences.

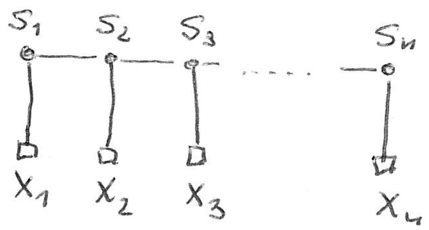
1c. Hidden Markov models on chains

Common situation in pattern recognition:

$X = (X_1, \dots, X_n)$ sequence of features (observable)
$S = (S_1, \dots, S_n)$ sequence of states (hidden)

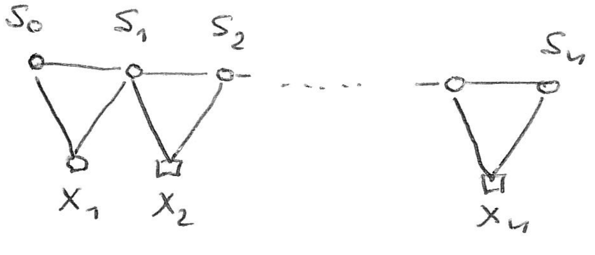
Hidden Markov model (HMM): a p.d. on pairs (x, s) s.t.

a) $$p(x,s) = \underbrace{\prod_{i=1}^{n} p(x_i | s_i)}_{p(x|s)} \cdot \underbrace{p(s_1) \cdot \prod_{i=2}^{n} p(s_i | s_{i-1})}_{p(s)\ \text{- Markov model}}$$



b) or, slightly more general

$$p(x,s) = p(s_0) \prod_{i=1}^{n} p(x_i, s_i | s_{i-1})$$



(stochastic regular language!)