I Markov Models on chains and acyclic graphs

1. Markov models on chains

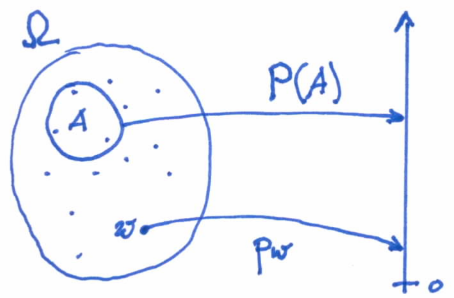
1A. Definitions and basic properties

- Sequence $s = (s_1, \ldots, s_n)$ of K-valued random variables $s_i \in K$, $i = 1, 2, \ldots, n$
- K is a finite set. We call its elements states
- $p(s) = p(s_1, \ldots s_n)$ denotes the joint probability distribution (p.d.) on K^n

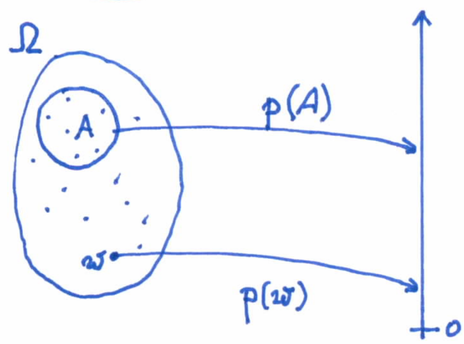
Agreement on notations:

a) Reminder: probability $\hat{=}$ a set function defined on a sample space Ω s.t. ....

often



here



p_ω – probability mass

P – probability

$p_\omega \hat{=} P \hat{=} p(.)$

b) $p(s_i)$ – marginal distribution of the variable s_i, i.e.

$$p(s_i = k) = \sum_{s:\, s_i = k} p(s_1, \ldots, s_n)$$

Similar notation for conditional prob's, e.g.

$$p(s_i \mid s_j) \quad \text{or} \quad p(s_i = k \mid s_j = \ell)$$

Let $p: K^n \to \mathbb{R}_+$ define an arbitrary p.d. W.l.o.g. we may write

$$p(s_1,\dots,s_n) = p(s_n | s_1,\dots,s_{n-1}) \cdot p(s_1,\dots,s_{n-1})$$
$$= \dots$$
$$= p(s_n | s_1,\dots,s_{n-1}) \cdot p(s_{n-1} | s_1,\dots,s_{n-2}) \cdot \dots \cdot p(s_1)$$

<u>Definition 1a</u> A p.d. on K^n is a Markov chain if

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

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

<u>Definition 1b</u> 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 $\forall\, s \in K^n$, where $g_i: K^2 \to \mathbb{R}_+$ are some (non-negative) functions.

<u>Equivalence</u>:

a) $\to$ b) trivial

b) $\to$ a) recursively apply

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

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

Therefore

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

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 unit, independently of the past, a particle is selected at random and moved to the other container. Let s_i denote the number of particles in the first container at time unit i. Then we have

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

Example 2 (Random walk on a graph)

Consider a random walk on an undirected graph (V, E).

- $K = V$ states
- $s_t \in V$ denotes the position of the walker at time unit t.
- $p(s_1)$ is 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 $w_{ij} \geq 0$ fulfil $\sum_{i \in \mathcal{N}(j)}' w_{ij} = 1 \;\; \forall\, j \in V$.

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

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

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

We know that

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

Consider $p(s_i = k)$, $k \in K$ as a vector $\vec{\pi}_i \in \mathbb{R}_+^K$ and $p(s_i = k \mid s_{i-1} = k')$, $k, k' \in K$ as a $K \times K$ matrix $\mathbb{P}$.

The previous eq. reads

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

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

We call it <u>stationary</u> p.d.

<u>Definition 2</u> A homogeneous Markov chain is <u>irreducible</u> if for each pair $k, k' \in K$ there is an $m > 0$ s.t. $\mathbb{P}^m_{kk'} > 0$.

I.e., there is a non-zero probability to reach state k starting from state k' (after m transitions). □

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

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

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 $0 < h < 1$. $\square$

Remark 1 (Infinite Markov chains)

- Consider infinite sequences $s = (s_1, s_2, \dots\dots)$, $s_i \in K$. $K^{\mathbb{N}}$ is uncountably infinite. Any probability on $K^{\mathbb{N}}$ will assign zero probability to (almost) each 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

$$(k_1, k_2, \dots, k_n) \longmapsto \{ s \in K^{\mathbb{N}} \mid s_1 = k_1, \dots, s_n = k_n \}$$

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

C. Hidden Markov models on chains

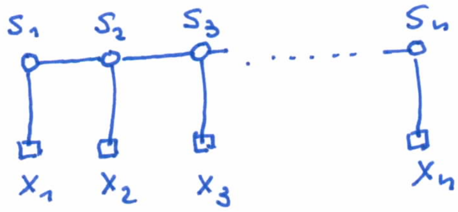
Common situation in pattern recognition:

$x = (x_1, \dots, x_n)$ sequence of features (observable)

$s = (s_1, \dots, s_n)$ sequence of states (hidden)

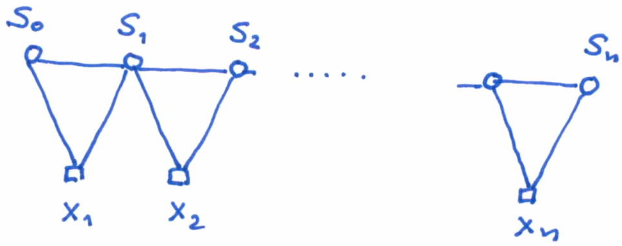
Hidden Markov model (HMM): a p.d. on pairs (x, s) s.t.

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



b) or, slightly more general

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



(stochastic regular language!)