Ch I Markov models on chains and acyclic graphs

1. Markov models on chains

1A Definitions & 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
* $p(s) = p(s_1, \ldots, s_n)$ is a joint probability distr. on $K^n$

Wlog. we can write

\[ p(s_1, \ldots, s_n) = p(s_1 | s_1, \ldots, s_{n-1}) p(s_2, \ldots, s_{n-1}) = \cdots = p(s_n | s_n, \ldots, s_1) p(s_1) \]

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

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

holds $\forall s \in K^n$

**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 $\forall s \in K_1^n$ where $g_i : K \to \mathbb{R}_+$ are some functions

**Equivalence:**

a) $\Rightarrow$ b) trivial

b) $\Rightarrow$ a) recursively apply the following step

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

$\Rightarrow$

\[ g_n(s_{n-1}, s_n) = p(s_n | s_{n-1}) b_{n-1}(s_{n-1}) \text{ with same } b_{n-1} \]
Therefore, we have
\[
p(s_{1}, s_{n}) = \left[ \prod_{i=2}^{n-1} p(s_{i-1}, s_{i}) \right] \cdot \frac{b_{n-1}(s_{n-1}) \cdot \rho(s_{n-1})}{\rho(s_{n-1}, s_{n-1})}
\]

Another useful formula
\[
p(s_{1}, s_{n}) = \frac{p(s_{1}, s_{2}) \cdot p(s_{2}, s_{3}) \cdots \cdot p(s_{n-1}, s_{n})}{\rho(s_{2}, \rho(s_{3}) \cdots \rho(s_{n-1})}
\]

Example 1 (Ehrenfest model)
Consider \( N \) particles in two containers. At each discrete time \( t = 1, 2, \ldots \), independently from 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
\[
p(S_{t} = k | S_{t-1} = \ell) = \begin{cases} \frac{N-k}{N} & \text{if } k = \ell + 1 \\ \frac{k}{N} & \text{if } k = \ell - 1 \\ 0 & \text{otherwise} \end{cases}
\]

Q: How does \( p(S_{t} = k), k = 1, \ldots, N \) behave for \( t \to \infty \)?

Example 2 (Random walk on a graph)
Consider a random walk on an undirected graph \( G = (V, E) \)
\[\begin{align*}
\bullet & \text{ } K = V \text{ } \text{states}, \quad S_{t} \in V \text{ } \text{position of the walker at time } t \\
\bullet & \text{ } p(s_{1}) \text{ some p.d. for the start vertex} \\
\bullet & \text{ } p(s_{t} = i | s_{t-1} = j) = \begin{cases} W_{ij} & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}
\end{align*}\)
where the \( W_{ij} \) fulfill \( \sum_{i \in \mathcal{V}(j)} W_{ij} = 1 \) \( \forall j \in V \)
8. Homogeneous Markov chains, stationary distributions

**Definition 2** A Markov chain is **homogeneous** if its conditional
prob's \( p(s_i; s_{i-1}) \) do not depend on the position \( i \), i.e.
\[
p(s_i = k' | s_{i-1} = k) = q(k, k') \quad \forall i = 1, \ldots, n.
\]

We know that
\[
p(s_i = k) = \sum_{k' \in \mathcal{K}} p(s_i = k | s_{i-1} = k') p(s_{i-1} = k').
\]

Consider \( p(s_i = k), k \in \mathcal{K} \) as components of a vector \( \pi_i \in \mathbb{R}_+^\mathcal{K} \)
and \( p(s_i = k | s_{i-1} = k'), k, k' \in \mathcal{K} \) as elements of a \( \mathcal{K} \times \mathcal{K} \) matrix \( P \).
Then the previous eq. reads
\[
\pi_i = P \pi_{i-1}
\]
and more general, we have \( \pi_i = P^{i-1} \pi_1 \).

It may happen that there \( \exists \) a p.d. \( \pi^* \) on \( \mathcal{K} \) s.t. \( P \pi^* = \pi^* \).
We call it a **stationary** p.d. of \( P \).

**Definition 3** A homogeneous Markov chain is **irreducible**
if for each pair of states \( k, k' \) there is an \( m \geq 0 \) s.t.
\( P^m_{k, k'} > 0 \). I.e. there is a non-zero probability to reach
state \( k \) starting from state \( k' \) (after \( m \) transitions).

A somewhat stronger condition ensures the existence &
uniqueness of a stationary distribution and convergence
to it.

**Theorem 1** (w/o proof) If for some \( m \geq 0 \) all elements
of the matrix \( P^m \) are strictly positive, then the Markov
chain has a **unique** stationary distribution \( \pi^* \), which
is a fixedpoint
\[
P^m \pi \xrightarrow{n \to \infty} \pi^* \quad \forall \pi
\]
Moreover
\[ P^n = \Pi^\ast \otimes \mathbf{E} + \mathbf{E}(h), \]
where \( \mathbf{E} = (e_{i,j}) \) and \( E_{kk'}(n) = O(h^n) \) with some \( 0 < h < 1 \).

**Definition 4.** A Markov chain satisfies the detailed balance condition if it has a stationary distribution \( \Pi \in \mathbb{R}^k_+ \) s.t.
\[ \pi(s_i | s_{i-1}) \pi(s_{i-1}) = \pi(s_{i-1} | s_i) \pi(s_i). \]
This means that the reverse Markov chain has the same transition probability matrix as the forward chain.

**C. Hidden Markov models on chains**

Common models in pattern recognition:
- \( X = (x_1, \ldots, x_n) \) sequence of features (observable)
- \( S = (s_1, \ldots, s_n) \) sequence of states (hidden)

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

- \( p(x, s) = \prod_{i=1}^{\infty} p(x_i | s_i) \cdot p(s_i) \cdot \prod_{i=2}^{\infty} p(s_i | s_{i-1}) \)

\[ p(x | s) \]

\[ p(s) \] - Markov model

6) or slightly more general

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

\[ p(x, s) \]

\[ p(s) \] - Markov model

**Remark.** This describes a stochastic regular language.