Markov Chains (MC)

- $\{X_n\}_{n\in\mathbb{N}_0}$ is called MC with state space $\mathcal{X}=\{x_1,\dots,x_m\}$ if
$$P(X_n=s_n \mid X_{n-1}=s_{n-1},\dots,X_0=s_0)=P(X_n=s_n \mid X_{n-1}=s_{n-1})$$
$$\forall s_i\in\mathcal{X},\ n\in\mathbb{N}.$$
- It is called <u>homogeneous</u>, if the transition probabilities $P(X_n=s_n \mid X_{n-1}=s_{n-1})$ are independent of n.
- $p(0)=(p_1(0),\dots,p_m(0))\sim X_0$ is called <u>initial</u> distribution.
- $\Pi=(p_{ij})_{1\le i,j\le m}=\left(P(X_n=j \mid X_{n-1}=i)\right)_{i,j=1,\dots,m}$ is called <u>transition matrix.</u>
- $p=(p_1,\dots,p_m)$ is called <u>stationary</u> if $p\Pi=p$.

<u>Lemma 2.3.6</u> Let $X=\{X_n\}_{n\in\mathbb{N}_0}$ be a stationary homogeneous MC. Then
$$H_\infty(X)=-\sum_{i,j}p_i(0)\,p_{ij}\log p_{ij}.\ \lrcorner$$

<u>Remark.</u> A homogeneous MC is stationary if
$$p(0)\,\Pi=p(0),\ \text{i.e.,}$$
if the initial distr. is a stationary distr. $\lrcorner$

Proof of Lem. 2.3.6

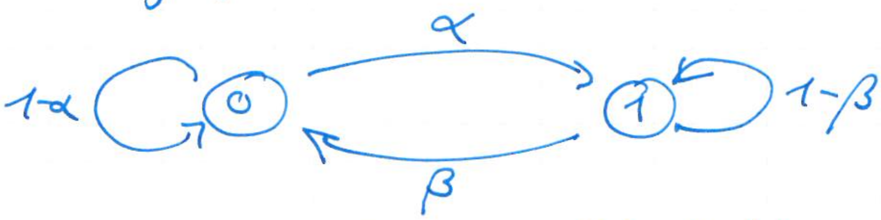
$$H_\infty(X) = \lim_{n\to\infty} H(X_n \mid X_{n-1}, \ldots, X_1)$$
$$= \lim_{n\to\infty} H(X_1 \mid X_0)$$
$$= -\sum_{i,j} p_i(0)\, p_{ij} \log p_{ij}. \quad \blacksquare$$

Example 2.3.7. (2-state MC)

Two states $\mathcal{X} = \{0,1\}$

Transition probabilities $\Pi = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$, $0 \le \alpha, \beta \le 1$

Transition graph



Compute the stationary distribution $p^* = (p_0, p_1)$,

solve $\quad p\Pi = p$.

Solution: $\quad p^* = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right) \qquad$ (Exercise)

Choose $p(0) = p^*$. Then $X = \{X_n\}_{n\in\mathbb{N}_0}$ is a stationary MC with

$$H(X_n) = H\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right) \quad \text{for all } n\in\mathbb{N}_0$$

However:

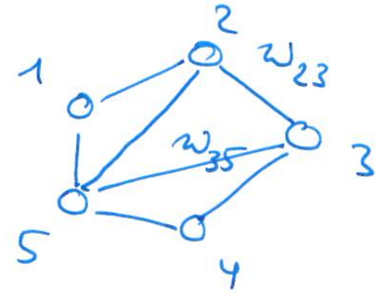
$$H_\infty(X) = H(X_1 \mid X_0) =$$
$$= \frac{\beta}{\alpha+\beta} H(\alpha, 1-\alpha) + \frac{\alpha}{\alpha+\beta} H(\beta, 1-\beta)$$

Exercise: Show that $H_\infty(X) \le H(X_n) = H(X_1)$

Example 2.3.8. (Random Walk on a weighted graph)

Consider an undirected weighted graph

Example



5 nodes
Edge between i and j has weight w_{ij}

Nodes: $\{1, \dots, m\}$

Edges with weights w_{ij}, $i<j = 1, \dots, m$, $w_{ji} = w_{ij}$

"no edges" means $w_{ij} = 0$.

Random walk on the graph $X = \{X_n\}_{n \in \mathbb{N}_0}$ is a MC with support $\mathcal{X} = \{1, \dots, m\}$ and

$$P(X_{n+1} = j \mid X_n = i) = \frac{w_{ij}}{\sum_{k=1}^{m} w_{ik}} = p_{ij}, \quad 1 \le i, j \le m.$$

Stationary distribution: (we guess it and then prove it is actually the one)

$$p_i^* = \frac{\sum_{j=1}^{m} w_{ij}}{\sum_{i,j} w_{ij}} = \frac{w_i}{w}$$

$$p^* = (p_1^*, \dots, p_m^*)$$

Set $\Pi = (p_{ij})_{1 \le i,j \le m}$

$$w_i := \sum_j w_{ij} \,, \quad w = \sum_{i,j} w_{ij}$$

$$(p^* \Pi)_j = \sum_i \frac{\cancel{\sum_\ell w_{i\ell}}}{w} \frac{w_{ij}}{\cancel{\sum_\ell w_{i\ell}}}$$

$$= \frac{1}{w} \sum_i w_{ij} \underset{\substack{\uparrow \\ \text{Symmetry}}}{=} \frac{1}{w} \sum_i w_{ji} = p_j^* .$$

$\boxed{\equiv}$

Assume the random walk starts at time 0 with the stat. distribution. $p_i(0) = p_i^*$, $i = 1, \dots, m$.
Then $\{X_n\}_{n \in \mathbb{N}_0}$ is a stationary MC and

$$H_\infty(X) = H(X_1 | X_0)$$

$$= - \sum_i p_i^* \sum_j p_{ij} \cdot \log p_{ij}$$

$$= - \sum_i \frac{w_i}{w} \sum_j \frac{w_{ij}}{w_i} \log \frac{w_{ij}}{w_i}$$

$$= - \sum_{i,j} \frac{w_{ij}}{w} \log \frac{w_{ij}}{w_i}$$

$$= - \sum_{i,j} \frac{w_{ij}}{w} \log \frac{w_{ij}}{w} + \sum_{i,j} \frac{w_{ij}}{w} \log \frac{w_i}{w}$$

$$\cancel{= -H\left(\left(\frac{w_{ij}}{w} \log w_{ij}\right.\right.}$$

$$= H\left(\left(\frac{w_{ij}}{w}\right)_{i,j}\right) - H\left(\left(\frac{w_i}{w}\right)_i\right)$$

If all edges have equal weight, then

$$p_i = \frac{E_i}{2E} \,, \quad E_i = \text{no of edges emanating from node } i$$

$$E = \text{total no of edges}$$

In this case

$$H_\infty(X) = \log(2E) - H\left(\frac{E_1}{2E}, \dots, \frac{E_m}{2E}\right)$$

$H_\infty(X)$ depends only on the entropy of the stationary distr. and the total no. of edges. $\lrcorner$

2.4. Asymptotic Equipartition Property (AEP)

In information theory, the AEP is the analog of the law of large numbers (LLN).

Let $\{X_i\}_{i\in\mathbb{N}}$ be i.i.d. r.v.s, $X_i \sim X$.

LLN: $\frac{1}{n}\sum_{i=1}^{n} X_i \to E(X)$ a.e. (& in probability) $(n\to\infty)$

AEP: $(X_1,\dots,X_n)$ with joint pmf $p^{(n)}(x_1,\dots,x_n)$. Then

$$-\frac{1}{n}\log p^{(n)}(X_1,\dots,X_n) \text{ "close to" } H(X) \quad (n\to\infty)$$

equivalently

$$p^{(n)}(X_1,\dots,X_n) \text{ "close to" } 2^{-nH(X)} \quad (n\to\infty)$$

"Close to" must be made precise.

Def. 2.4.1. A sequence of r.v. X_n is said to converge to a r.v. X

(i) in probability if $\forall \varepsilon>0$: $P(|X_n - X|>\varepsilon)\to 0$ $(n\to\infty)$

(ii) in mean square if $E[(X_n-X)^2]\to 0$ $(n\to\infty)$

(iii) with probability 1 (or almost surely / everywhere) if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1 \quad \lrcorner$$

$$X_n, X : (\Omega, \mathcal{A}, P) \to \mathbb{R}^1$$

$$P\left(\lim_{n\to\infty} X_n = X\right) = P\left(\{\omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1$$

Theorem 2.4.2. (AEP)

Let $\{X_n\}$ be i.i.d. discrete r.v.s, $X_i \sim X$ with support $\mathcal{X} = \{x_1, \dots, x_m\}$. $(X_1, \dots, X_n)$ with joint pmf $p^{(n)}(x_1, \dots, x_n)$. Then

$$-\frac{1}{n} \log p^{(n)}(X_1, \dots, X_n) \xrightarrow[(n\to\infty)]{} H(X) \text{ in probability.}$$

Proof. $Y_i = \log p(X_i)$ are also i.i.d. By LLN

$$-\frac{1}{n} \log p^{(n)}(X_1, \dots, X_n) = -\frac{1}{n} \sum_{i=1}^{n} \log p(X_i)$$
$$\xrightarrow[(n\to\infty)]{} E(-\log p(X)) = H(X).$$

with convergence in probability. $\boxtimes$

Def. 2.4.3. Situation as in Th. 2.4.2.

$$A_\varepsilon^{(n)} = \left\{ (x_1, \dots, x_n) \in \mathcal{X}^n \,\middle|\, 2^{-n(H(X)+\varepsilon)} \le p^{(n)}(x_1, \dots, x_n) \le 2^{-n(H(X)-\varepsilon)} \right\}$$

is called typical set w.r.t. ε and p.

Th. 2.4.4.

a) If $(x_1, \dots, x_n) \in A_\varepsilon^{(n)}$ then

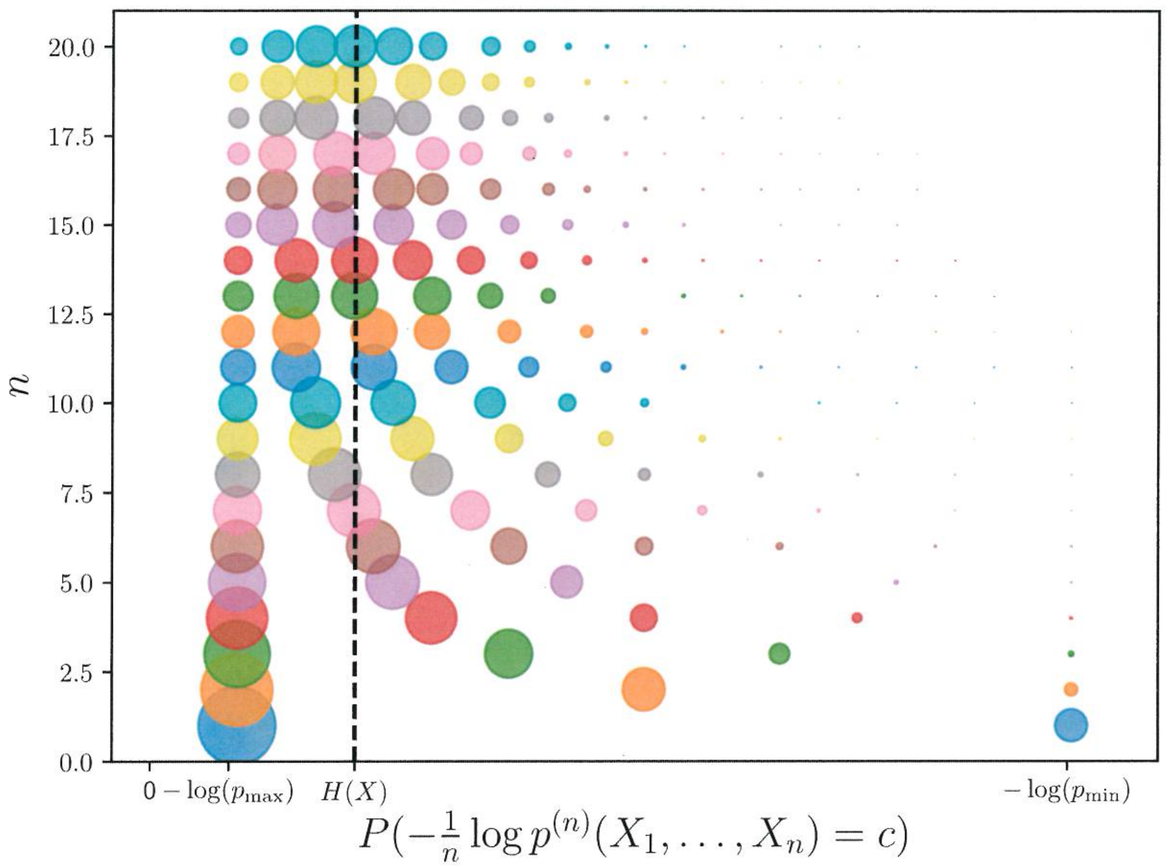
$$H(X) - \varepsilon \le -\frac{1}{n} \log p^{(n)}(x_1, \dots, x_n) \le H(X) + \varepsilon$$

b) $P(A_\varepsilon^{(n)}) > 1 - \varepsilon$ for sufficiently large n.

c) $|A_\varepsilon^{(n)}| \le 2^{n(H(X)+\varepsilon)}$ ($|\cdot|$: cardinality)

d) $|A_\varepsilon^{(n)}| \ge (1-\varepsilon)\, 2^{n(H(X)-\varepsilon)}$ for sufficiently large n.

$p = (0.85, 0.15)$



20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
n
0
$-\log(p_{\max})$
$H(X)$
$-\log(p_{\min})$
$P(-\frac{1}{n}\log p^{(n)}(X_1, \ldots, X_n) = c)$

$p = (0.6, 0.3, 0.1)$