Markov Chains

A stochastic process $\{X_i\}_{i\in I}$ is an indexed family of random variables (not necessarily independent). Usually, we will be dealing with stoc processes indexed by $\mathbb{N}$, i.e. sequences of random variables.

A stochastic process is **stationary** if it is invariant with respect to shifts in the time index, that is:

$$P(X_1=a_1,\ldots,X_n=a_n)=P(X_{k+1}=a_1,\ldots,X_{k+n}=a_n)$$

for all $n,k\in\mathbb{N}$, and $a_i\in\mathcal{X}$.

P1 Prove that for an arbitrary st. pr. $\{X_i\}_{i\in\mathbb{N}}$

$$H(X_{n+1}\mid X_1,\ldots,X_n)\le H(X_n\mid X_1,\ldots,X_{n-1})$$

Sol:

$$H(X_{n+1}\mid X_1,\ldots,X_n)\le H(X_{n+1}\mid X_2,\ldots,X_n)$$

$$H(X_2,\ldots,X_{n+1})=H(X_1,\ldots,X_n)=H(X_1,\ldots,X_{n-1})+H(X_n\mid X_1\ldots X_{n-1})$$

$$\parallel$$

$$H(X_2,\ldots,X_n)+H(X_{n+1}\mid X_2,\ldots X_n)$$

$$\parallel$$

$$H(X_1,\ldots X_{n-1})$$

$\mathcal{S} = \{X_i\}_{i \in \mathbb{N}}$

For an arbitrary stochastic process we define the entropy of the proces as:

$$H(\mathcal{S}) = \lim_{n \to +\infty} \frac{1}{n} H(X_1, \dots, X_n),$$

when the limit exists.

We can also define the entropy rate of the process as:

$$H'(\mathcal{S}) = \lim_{n \to +\infty} H(X_n \mid X_1, \dots, X_{n-1})$$

P2 Prove that for a stationary stochastic process the entropy exists an that

$$H(\mathcal{S}) = H(\mathcal{S}').$$

Sol: By the chain rule for entropy we have

$$\frac{1}{n} H(X_1, \dots, X_n) = \frac{1}{n} \sum_{k=1}^{n} H(X_k \mid X_1, \dots, X_{k-1}).$$

Now from P1, we know that for $a_k = H(X_k \mid X_1 \dots X_{k-1})$ $\{a_k\}$ is a sequence of non increasing real numbers, bounded from below with 0, so

$\lim_{k\to+\infty} a_k = \lim_{k\to+\infty} H(X_k | X_1 \dots X_{k-1})$ exists (and is between (0 and $H(X_1)$).

We need to prove that if $a_k \to a$, then $\frac{1}{k}\sum_{i=1}^{k} a_i \xrightarrow[k\to+\infty]{} a$. (Cesaro mean th.)

For $\varepsilon > 0$, let N_ε be such that

$|a_k - a| < \frac{\varepsilon}{2}$, then:

$$\left| \frac{1}{k}\sum_{i=1}^{k} a_i - a \right| \leq \sum_{i=1}^{N_\varepsilon} \frac{|a_i - a|}{k} + \overbrace{\left(\frac{k - N_\varepsilon}{k}\right)}^{\leq 1} \frac{\varepsilon}{2} < \varepsilon$$

for k large enough (because N_ε is fixed),

so $\lim_{k\to+\infty} \frac{1}{k}\sum_{i=1}^{k} a_i \to a$.

P3 For a stationary st. pr. $X_1, X_2, \dots X_n, \dots$ show that:

a) $$\frac{H(X_1, \dots, X_n)}{n} \leq \frac{H(X_1, \dots, X_{n-1})}{n-1}$$

b) $$\frac{H(X_1, \dots, X_n)}{n} \geq H(X_n | X_1, \dots, X_{n-1}).$$

A stochastic process is said to be a Markov chain if for all $n = 1, 2, \dots$

$$P(X_{n+1} = a_{n+1} \mid X_1 = a_1, \dots, X_n = a_n) = P(X_{n+1} = a_{n+1} \mid X_n = a_n)$$

for all $a_i \in A$.

A Markov chain is said to be time invariant if

$$P(X_{n+1} = a \mid X_n = b) = P(X_2 = a \mid X_1 = b)$$

for all $a, b \in A$.

When M.C. is time-invariant we can form a prob. transition matrix of the M.C. as

$$P = [P_{ij}] \quad , \text{where}$$

$$P_{ij} = P(X_{n+1} = x_j \mid X_n = x_i) \quad .$$

A distribution on the possible states ν such that the distribution at time $n+1$ is the same as the one at time n is called stationary distribution.

P4 Prove that for a stationary Markov chain $\mathcal{C}$ the entropy rate is

$$H(\mathcal{C}) = H(X_2 | X_1)$$

Sol:

$$\frac{H(X_1, \dots, X_n)}{n} =$$

$$= \frac{\sum_{k=1}^{n} H(X_k | X_1 \dots X_{k-1})}{n}$$

$$= \frac{\sum H(X_k | X_{k-1})}{n}$$

$$= \frac{n \cdot H(X_2 | X_1)}{n}$$

$$= H(X_2 | X_1)$$

(Here the conditional entropy is calculated using the stationary distribution)

P5 A Professor gives various versions of an exam, with students taking the exam one after the other. Each version can be difficult, normal or easy.
After a difficult exam the next one will be diff. with pr. 0.2, normal 0.5 and easy 0.3. After normal and easy: 0.5, 0.25 and 0.25 respectively.

a) What is the steady state dist. for this M.C.
b) Find the entropy rate for this M.C.

$$P = [P_{ij}] = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \end{bmatrix}.$$

$(H_\infty = 1.4944)$

P6

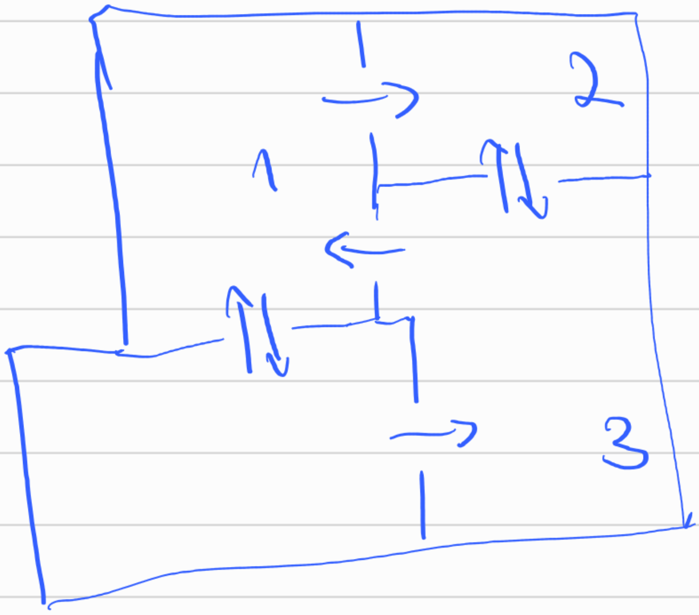
Let the prob. dist. matrix of a M.C. be given by:

$$\Pi = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0.1 & 0 & 0.9 \end{pmatrix}$$

a) Calculate the stationary distr. of the process
b) Find the entropy rate.

<u>Sol:</u> $\left(\frac{3}{11}, \frac{2}{11}, \frac{6}{11}\right)$ and $H_\infty = 0.613$

P7 In the figure below the plan for a house is given. When a cat is in a room it can stay or leave using outgoing doors. It chooses between the options with equal prob. There are ants in room 3, and the cat cannot stay there.



a) Find the steady state distr.
b) Find the entropy rate.

Sol: $\bar{u} = \left(\frac{6}{25}, \frac{2}{5}, \frac{6}{25}, \frac{3}{25} \right)$

$H_p = 1.21$

P8

Doubly stochastic matrices. An $n \times n$ matrix $P = [P_{ij}]$ is said to be *doubly stochastic* if $P_{ij} \geq 0$ and $\sum_j P_{ij} = 1$ for all i and $\sum_i P_{ij} = 1$ for all j. An $n \times n$ matrix P is said to be a *permutation* matrix if it is doubly stochastic and there is precisely one $P_{ij} = 1$ in each row and each column. It can be shown that every doubly stochastic matrix can be written as the convex combination of permutation matrices.

(a) Let $\mathbf{a}^t = (a_1, a_2, \ldots, a_n)$, $a_i \geq 0$, $\sum a_i = 1$, be a probability vector. Let $\mathbf{b} = \mathbf{a}P$, where P is doubly stochastic. Show that $\mathbf{b}$ is a probability vector and that $H(b_1, b_2, \ldots, b_n) \geq H(a_1, a_2, \ldots, a_n)$. Thus, stochastic mixing increases entropy.

(b) Show that a stationary distribution μ for a doubly stochastic matrix P is the uniform distribution.

(c) Conversely, prove that if the uniform distribution is a stationary distribution for a Markov transition matrix P, then P is doubly stochastic.

Sol: a) From $\bar{b} = \bar{a}P$, we have:

$$b_j = \sum_i a_i P_{ij} \quad , \text{ so}$$

$$H(\bar{b}) = \sum_j b_j \log_2 \frac{1}{b_j} = \sum_j \sum_i a_i P_{ij} \log \frac{1}{b_j}$$

$$= \sum_i a_i \left(\sum_j P_{ij} \log \frac{1}{b_j} \right)$$

$$H(b) - H(a) = \sum_k \Big(\underbrace{b_k \log \frac{1}{b_k}}_{\sum_i a_i P_{ik} \log \frac{1}{b_k}} + a_k \log a_k \Big)$$

$$\sum_i a_i P_{ik} \log \frac{1}{b_k} + a_k \log a_k$$

$$\sum_k \sum_i a_i P_{ik} \log \frac{1}{b_k} + \sum_i \sum_k a_k P_{ki} \log a_k$$

$$= \sum_{i,k} a_i P_{ik} \log \frac{a_i}{b_k} = \sum_{i,k} a_i P_{ik} \log \frac{a_i P_{ik}}{\left(\sum_t a_t P_{tk}\right) P_{ik}}$$

$$\geq 0 \quad (\text{By Gibbs inequality})$$

c) From:

$$\alpha P_{ij} = \alpha$$

$$\sum_i \frac{1}{n} P_{ij} = \frac{1}{n} \Rightarrow \sum_i P_{ij} = 1$$

So $[P_{ij}]$ is doubly stoc.

c) $$\left(\frac{1}{n}, \dots \frac{1}{n}\right) P = \left(\frac{1}{n} \sum_j P_{j1} \dots \frac{1}{n} \sum_i P_{jn}\right)$$

$$= \left(\frac{1}{n}, \dots, \frac{1}{n}\right).$$

2. Let the probability transition matrix of a Markov process be given by,

$$\Pi = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0.1 & 0 & 0.9 \end{pmatrix}$$

That is, there are three states S_1, S_2, S_3 and the entries $\pi_{i,j}$ denotes the probability of going to state S_j from state S_i.

(a) Calculate the stationary distribution of this process.

(b) Find the entropy rate H_∞.

(c) Denote the states S_1, S_2, S_3 by A, B, C respectively. Consider the extended alphabet by looking at two states so that the alphabet now consists of AA, AB, AC, BB, BC, CC. What is the average number of bits per symbol in this case if Huffman coding is used? Hint: Do not forget the stationary distribution of the symbols A, B, C, e.g. $Pb(AB)$ means that you are in state A and going to state B.

Solution:

a) Using $\mu\Pi = \mu$, where $\mu = (\mu_1, \mu_2, \mu_3)$, and $\mu_1 + \mu_2 + \mu_3 = 1$ one gets

$$(\mu_1, \mu_2, \mu_3) = \left(\frac{3}{11}, \frac{2}{11}, \frac{6}{11}\right).$$

b) Assuming the process is ergodic thus applying the formula $H_\infty = \sum_i P(S_i)H(S_i)$, where S_i denote the states of the Markov process.

The entropy rate $H_\infty = \sum_i P(S_i)H(S_i)$ can be written as $\sum_i \mu_i \sum_j P_{i,j} \ln P_{i,j}$ which then gives

$$\begin{aligned} H_\infty = \; & (3/11)(-0.8\log_2 0.8 - 0.2\log_2 0.2) + (2/11)(-0.7\log_2 0.7 - 0.3\log_2 0.3) + \\ & + (6/11)(-0.9\log_2 0.9 - 0.1\log_2 0.1) = 0.613. \end{aligned}$$

c) The probabilities of the symbols are

$$\begin{aligned} Pb(AA) &= 0.8 \cdot 3/11 = 24/110 \\ Pb(AB) &= 0.2 \cdot 3/11 = 6/110 \\ Pb(BB) &= 0.7 \cdot 2/11 = 14/110 \\ Pb(BC) &= 0.3 \cdot 2/11 = 6/110 \\ Pb(CC) &= 0.9 \cdot 6/11 = 54/110 \\ Pb(CA) &= 0.1 \cdot 6/11 = 6/110. \end{aligned}$$

Calculation of Average Bits per Extended Symbol:

Symbol	Calculation	Probability	Huffman Code	Length
CC	$\frac{6}{11} \times 0.9$	54/110	0	1
AA	$\frac{3}{11} \times 0.8$	24/110	10	2
BB	$\frac{2}{11} \times 0.7$	14/110	110	3
BC	$\frac{2}{11} \times 0.3$	6/110	1110	4
AB	$\frac{3}{11} \times 0.2$	6/110	11110	5
CA	$\frac{6}{11} \times 0.1$	6/110	11111	5

$$L_{avg} = \frac{1}{110}[54(1) + 24(2) + 14(3) + 6(4) + 6(5) + 6(5)] = \frac{208}{110} \approx 1.891 \text{ bits}$$

P-2 A Professor gives varying versions of an oral examination in assembly line fashion, with students taking the exam one after the other. Each version of the exam may be categorized as difficult, normal or easy. After a difficult exam, the next exam will be difficult with probability 0.2, will be normal with probability 0.5, and will be easy with probability 0.3. Similarly, after normal and easy exams, these probabilities are 0.5, 0.25 and 0.25 respectively.

1. What is the steady state distribution for this Markov chain?
2. Find the entropy rate H_∞ for this Markov chain (you might find the formula $H_\infty = \sum_i P(S_i)H(S_i)$ useful).

Solution: Let's call the states "hard exam", "medium exam", "easy exam" respectively 1, 2, 3.

From the information given to us in the statement of the problem we deduce that the transition probability matrix is:

$$M = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \end{bmatrix}$$

a) To get the steady state distribution we have to solve $(v_1, v_2, v_3)M = (v_1, v_2, v_3)$ and $v_1 + v_2 + v_3 = 1$. Multiplying the matrix we get the following system of linear equations:

$$\begin{aligned} v_1 + v_2 + v_3 &= 1 \\ \frac{1}{5}v_1 + \frac{1}{2}v_2 + \frac{1}{2}v_3 &= v_1 \\ \frac{1}{2}v_1 + \frac{1}{4}v_2 + \frac{1}{4}v_3 &= v_2 \\ \frac{3}{10}v_1 + \frac{1}{4}v_2 + \frac{1}{4}v_3 &= v_3 \end{aligned}$$

$$
\begin{aligned}
&1) \quad v_1 + v_2 + v_3 = 1 \\
&2) \quad \frac{1}{5}v_1 + \frac{1}{2}v_2 + \frac{1}{2}v_3 = v_1 \implies -\frac{4}{5}v_1 + \frac{1}{2}(v_2 + v_3) = 0 \\
&3) \quad \frac{1}{2}v_1 + \frac{1}{4}v_2 + \frac{1}{4}v_3 = v_2 \implies \frac{1}{2}v_1 - \frac{3}{4}v_2 + \frac{1}{4}v_3 = 0
\end{aligned}
$$

Substitute $\mathrm{v}_2 + v_3 = 1 - v_1$ into (2):
$-\frac{4}{5}v_1 + \frac{1}{2}(1 - v_1) = 0 \implies -8v_1 + 5 - 5v_1 = 0 \implies v_1 = \frac{5}{13}$

Substitute $\mathrm{v}_1 = \frac{5}{13}$ into (1) and (3):
$v_2 + v_3 = \frac{8}{13}$ and $\frac{5}{26} - \frac{3}{4}v_2 + \frac{1}{4}v_3 = 0$

Solving the resulting system yields:
$\mathrm{v}_1 = \frac{5}{13}, \quad v_2 = \frac{9}{26}, \quad v_3 = \frac{7}{26}$

The solution, and hence the steady state distribution, is $v_1 = \frac{5}{13}$, $v_2 = \frac{9}{26}$ and $v_3 = \frac{7}{26}$.

b) Using the formula given to us in the problem, we have:

$$
\begin{aligned}
H_\infty &= \sum_i P(S_i)H(S_i) \\
&= \frac{5}{13}\left(\frac{1}{5}\log 5 + \frac{1}{2}\log 2 + \frac{3}{10}\log\frac{10}{3}\right) + \frac{9}{26}\left(\frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4\right) + \frac{7}{26}\left(\frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\mathrm{lo}\right. \\
&= 1.4944.
\end{aligned}
$$