Markov sources and search algorithms

Universal source coding

Huffman coding is optimal, what is the problem?

In the previous coding schemes (Huffman and Shannon-Fano)it was assumed that

- The source statistics is known
- The source symbols are i.i.d.

Normally this is not the case.

How much can the source be compressed?
How can it be achieved?

Random processes

Definition (Random process)

A random process $\{X_i\}_{i=1}^{n}$ is a sequence of random variables. There can be an arbitrary dependence among the variables and the process is characterized by the joint probability function

$$P(X_1, X_2, \ldots, X_n = x_1, x_2, \ldots, x_n) = p(x_1, x_2, \ldots, x_n),\ n = 1, 2, \ldots$$

Definition (Stationary random process)

A random process is stationary if it is invariant in time,

$$P(X_1, \ldots, X_n = x_1, \ldots, x_n) = P(X_{q+1}, \ldots, X_{q+n} = x_1, \ldots, x_n)$$

for all time shifts q.

Entropy rate

Definition

The entropy rate of a random process is defined as

$$H_\infty(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1 X_2 \dots X_n)$$

Define the alternative entropy rate for a random process as

$$H(X|X^\infty) = \lim_{n\to\infty} H(X_n|X_1 X_2 \dots X_{n-1})$$

Theorem

The entropy rate and the alternative enropy rate are equivalent,

$$H_\infty(X) = H(X|X^\infty)$$

Special case – independent variables

We assume that X_i are I.I.D. and compute $H_\infty(X)$:

$$\begin{aligned} H_\infty(X) &= \lim_{n\to\infty} \frac{1}{n} H(X_1 \dots X_n) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_1 \dots X_{i-1}) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i) \\ &= \lim_{n\to\infty} \frac{1}{n} H(x) \sum_{i=1}^{n} 1 = H(X). \end{aligned}$$

Decreasing entropy chain

Considering a stationary (time-invariant) random process, it can be seen that:

$$H(X_n|X_1 \dots X_{n-1}) \leq H(X_n|X_2 \dots X_{n-1}) = H(X_{n-1}|X_1 \dots X_{n-2}),$$

where the last equality comes from stationary property (decreasing). Then

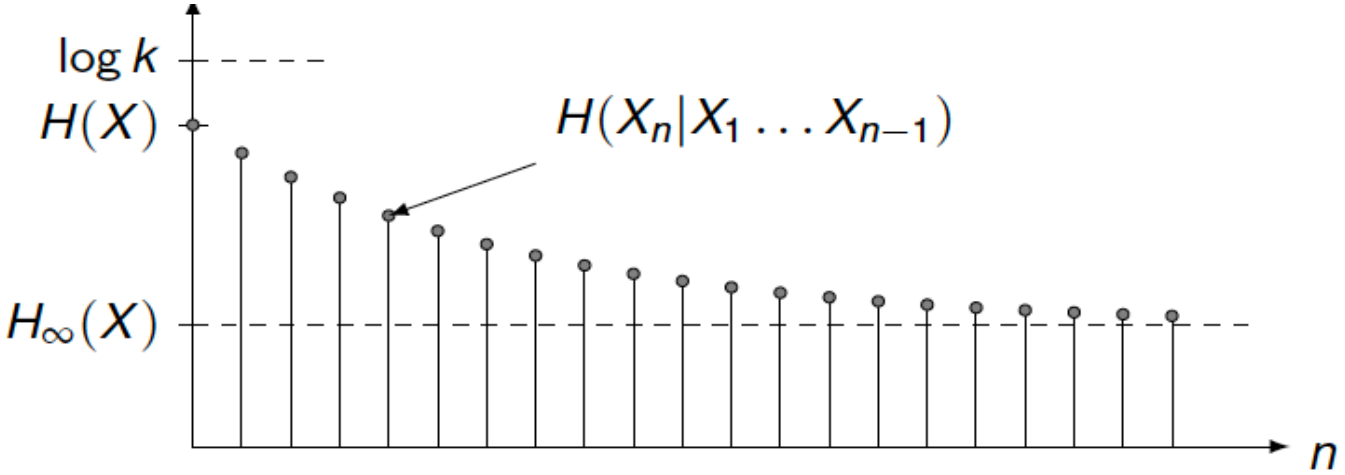
$$H(X_n|X_1 \dots X_{n-1}) \leq \cdots \leq H(X_2|X_1) \leq H(X_1) = H(X) \leq \log_2 |\mathcal{X}|.$$

Entropy rate

Theorem

For a stationary stochastic process the entropy rate is bounded by

$$0 \leq H_\infty(X) \leq H(X) \leq \log k$$



Source coding for random processes

Optimal coding of process

Let $\boldsymbol{X} = (X_1, \ldots, X_N)$ be a vector of N symbols from a random process. Use an optimal source code to encode the vector. Then

$$H(X_1 \ldots X_N) \leq L^{(N)} \leq H(X_1 \ldots X_N) + 1$$

which gives the average codeword length per symbol, $L = \frac{1}{N} L^{(N)}$,

$$\frac{1}{N} H(X_1 \ldots X_N) \leq L \leq \frac{1}{N} H(X_1 \ldots X_N) + \frac{1}{N}$$

In the limit as $N \to \infty$ the optimal codeword length per symbol becomes

$$\lim_{N \to \infty} L = H_\infty(X)$$

Markov chain

Definition (Markov chain)

A Markov chain, or Markov process, is a random process with unit memory,

$$P(x_n|x_1,\ldots,x_{n-1}) = P(x_n|x_{n-1}), \quad \text{for all } x_i$$

Definition (Stationary)

A Markov chain is stationary (time invariant) if the conditional probabilities are independent of the time,

$$P(X_n = x_a|X_{n-1} = x_b) = P(X_{n+\ell} = x_a|X_{n+\ell-1} = x_b)$$

for all relevant n, ℓ, x_a and x_b.

Markov chain

Theorem

For a Markov chain the joint probability function is

$$
\begin{aligned}
p(x_1, x_2, \ldots, x_n) &= \prod_{i=1}^{n} p(x_i | x_1, x_2, \ldots, x_{i-1}) \\
&= \prod_{i=1}^{n} p(x_i | x_{i-1}) \\
&= p(x_1) p(x_2 | x_1) p(x_3 | x_2) \cdots p(x_n | x_{n-1})
\end{aligned}
$$

Markov chain characterization

Definition

A Markov chain is characterized by

- A state transition matrix

$$P = [p(x_j|x_i)]_{i,j\in\{1,2,\ldots,k\}} = [p_{ij}]_{i,j\in\{1,2,\ldots,k\}}$$

  where $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$.

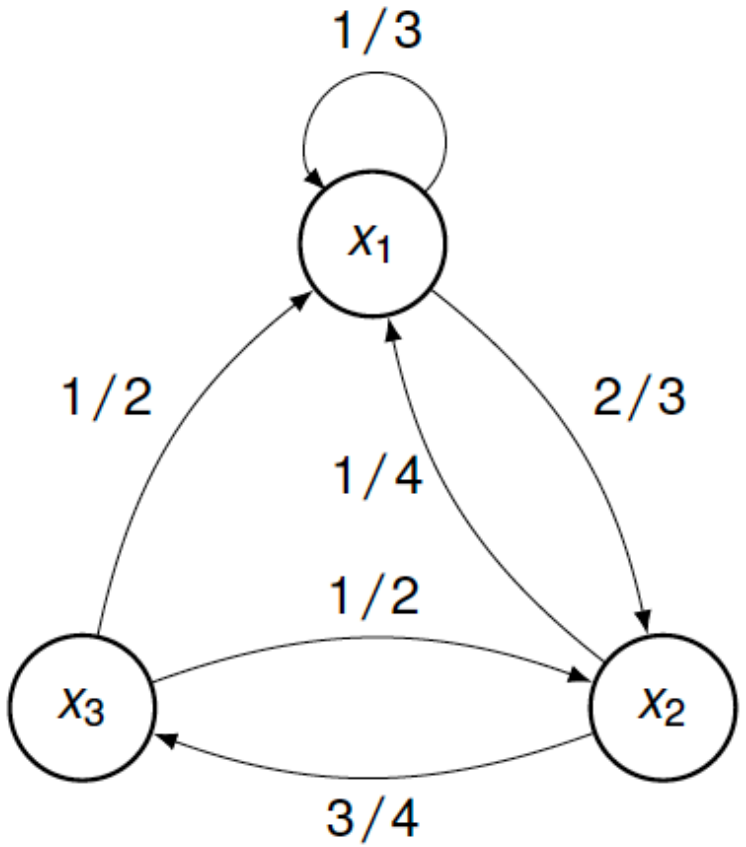
- A finite set of states

$$X \in \{x_1, x_2, \ldots, x_k\}$$

  where the state determines everything about the past.

The state transition graph describes the behaviour of the process

Example of Markov chain



The state transition matrix

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

The state space is

$$X \in \{x_1, x_2, x_3\}$$

Computing the next state

- Assume at $t = 0$ we have $Pb(X_0 = x_1) = Pb(X_0 = x_2) = Pb(X_0 = x_3) = 1/3$. What about $t = 1$?
- We compute:

$$Pb(X_1 = x_1) = \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{1}{2} = \frac{13}{36}$$

$$Pb(X_1 = x_2) = \frac{1}{3} \times \frac{2}{3} + \frac{1}{3} \times 0 + \frac{1}{3} \times \frac{1}{2} = \frac{14}{36}$$

$$Pb(X_1 = x_3) = 0 + \frac{1}{3} \times \frac{3}{4} = \frac{9}{36}$$

Of course, $\sum_{i=1}^{k} Pb(X_t = x_i) = 1$, in our case $k = 3$.

Markov chain – state distribution

Theorem

Given a Markov chain with k states, let the distribution for the states at time n be

$$\boldsymbol{\pi}^{(n)} = (\pi_1^{(n)}\ \pi_2^{(n)} \ldots \pi_k^{(n)})$$

Then

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)} P^n$$

where $\boldsymbol{\pi}^{(0)}$ is the initial distribution at time 0.

Proof

Denote $\pi_j^{(n)} = Pb(X_n = x_j)$ - probability at time $t = n$ we are in state x_j.
Then,

$$\pi_j^{(n)} = \pi_1^{(n-1)} p_{1j} + \pi_2^{(n-1)} p_{2j} + \cdots + \pi_k^{(n-1)} p_{kj},$$

for $j = 1, \ldots, k$.

This implies

$$\begin{aligned}
(\pi_1^{(n)}, \ldots, \pi_k^{(n)}) &= (\pi_1^{(n-1)}, \ldots, \pi_k^{(n-1)}) \times P \\
&= (\pi_1^{(n-2)}, \ldots, \pi_k^{(n-2)}) \times P \times P \\
&= \ldots \\
&= (\pi_1^{(0)}, \ldots, \pi_k^{()}) \times P^n.
\end{aligned}$$

Example- asymptotic distribution

$$P^2 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} = \begin{pmatrix} \frac{20}{72} & \frac{16}{72} & \frac{36}{72} \\ \frac{33}{72} & \frac{39}{72} & 0 \\ \frac{21}{72} & \frac{24}{72} & \frac{27}{72} \end{pmatrix}$$

$$P^4 = \begin{pmatrix} \frac{20}{72} & \frac{16}{72} & \frac{36}{72} \\ \frac{33}{72} & \frac{39}{72} & 0 \\ \frac{21}{72} & \frac{24}{72} & \frac{27}{72} \end{pmatrix} \begin{pmatrix} \frac{20}{72} & \frac{16}{72} & \frac{36}{72} \\ \frac{33}{72} & \frac{39}{72} & 0 \\ \frac{21}{72} & \frac{24}{72} & \frac{27}{72} \end{pmatrix} = \begin{pmatrix} \frac{1684}{5184} & \frac{1808}{5184} & \frac{1692}{5184} \\ \frac{1947}{5184} & \frac{2049}{5184} & \frac{1188}{5184} \\ \frac{1779}{5184} & \frac{1920}{5184} & \frac{1485}{5184} \end{pmatrix}$$

$$P^8 = \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad = \begin{pmatrix} 0.3485 & 0.3720 & 0.2794 \\ 0.3491 & 0.3721 & 0.2788 \\ 0.3489 & 0.3722 & 0.2789 \end{pmatrix}$$

Asymptotic distribution

- The entries in each column have approximately the same value !
- $p_{11}^{(8)} = p_{21}^{(8)} = p_{31}^{(8)}$ - coming to state x_1 is the same (does not depend on the initial state any more)
- In other words, for large n the probability p_{ij} does NOT depend on i.
- Alternatively, for large n, we get $P^{n+1} = P^n$!
- Think about $P \times P^n$ when P is of the form $P^n = \begin{pmatrix} A & B & C \\ A & B & C \\ A & B & C \end{pmatrix}$
- Make sense to define $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for $j = 1, \ldots, k$.

Markov chain – stationary distribution

Theorem

Let $\pi = (\pi_1 \dots \pi_k)$ be an asymptotic distribution of the state probabilities. Then

- $\sum_j \pi_j = 1$
- *π is a stationary distribution, i.e. $\pi P = \pi$*
- *π is a unique stationary disribution for the source.*

Proof

We show the second part (first one is obvious). The process being stationary we have $P \times P^n = P^{n+1} = P^n$. Then, multiplying with π we get

$$(\pi P) \times P^n = \pi \times P^n,$$

which implies that $\pi P = \pi$.

Which Markov chains are stationary ?

- **Answer: If and only if** there exists $N > 0$ such that P^N has a positive column, i.e. all elements in this column are > 0 !!

Computing stationary distribution

We need to compute the stationary distribution π using that $\sum_{i=1}^{k} \pi_i = 1$ and $\pi P = \pi$. The equation $\pi P = \pi$ gives:

$$\begin{aligned} 1/3 \times \pi_1 + 1/4 \times \pi_2 + 1/2 \times \pi_3 &= \pi_1 \\ 2/3 \times \pi_1 + 0 + 1/2 \times \pi_3 &= \pi_2 \\ 0 + 3/4 \times \pi_2 + 0 &= \pi_3 \end{aligned}$$

Solve and get : $\pi_1 = 0.3488$; $\pi_2 = 0.3721$; $\pi_3 = 0.2791$.

Entropy rate of Markov chain

Theorem

For a stationary Markov chain with stationary distribution π and transition matrix P, the entropy rate can be derived as

$$H_\infty(X) = \sum_i \pi_i H(X_2|X_1 = x_i)$$

where

$$H(X_2|X_1 = x_i) = -\sum_j p_{ij} \log p_{ij}$$

the entropy of row i in P.

Proof

For Markov sources we have $H(X_n|X_1 \dots X_{n-1}) = H(X_n|X_{n-1})$ and we use the fact that

$$H_\infty(X) = H(X|X^\infty) = \lim_{n\to\infty} H(X_n|X_1 \dots X_{n-1}).$$

Then,

$$\begin{aligned}
H_\infty(X) &= \lim_{n\to\infty} H(X_n|X_1 \dots X_{n-1}) = \lim_{n\to\infty} H(X_n|X_{n-1}) = H(X_2|X_1) \\
&= \sum_{i,j} P(X_1 = x_i, X_2 = x_j) \log_2 P(X_2 = x_j|X_1 = x_i) \\
&= \sum_i P(X_1 = x_i) \sum_j P(X_2 = x_j|X_1 = x_i) \log_2 P(X_2 = x_j|X_1 = x_i) \\
&= \sum_i H(X_2|X_1 = x_i) P(X_1 = x_i),
\end{aligned}$$

where $H(X_2|X_1 = x_i) = \sum_j P(X_2 = x_j|X_1 = x_i) \log_2 P(X_2 = x_j|X_1 = x_i)$.

Entropy rate - example

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

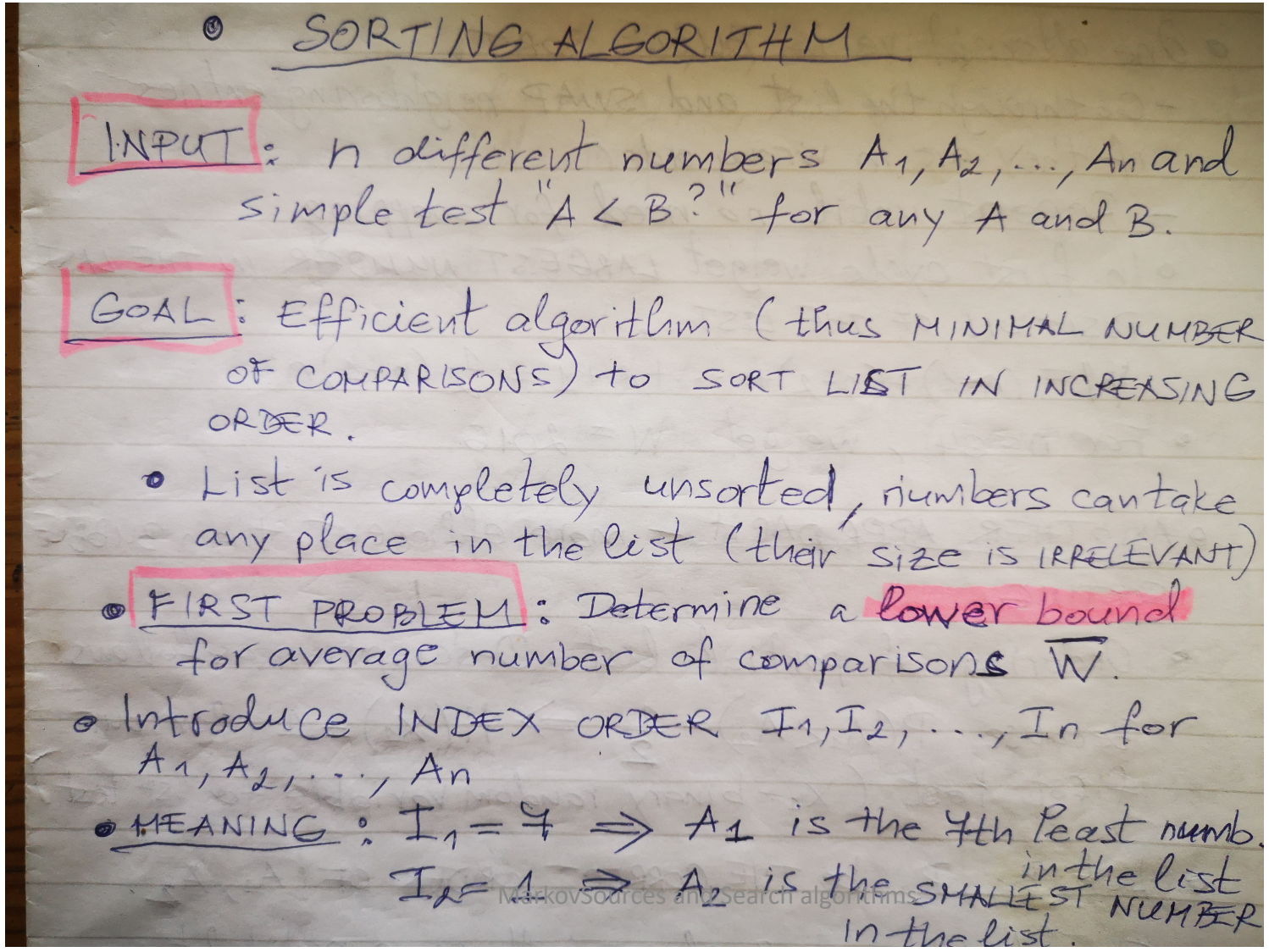
Entropy per row:

$$H(X_2|X_1 = x_1) = h(\tfrac{1}{3})$$
$$H(X_2|X_1 = x_2) = h(\tfrac{1}{4})$$
$$H(X_2|X_1 = x_3) = h(\tfrac{1}{2}) = 1$$

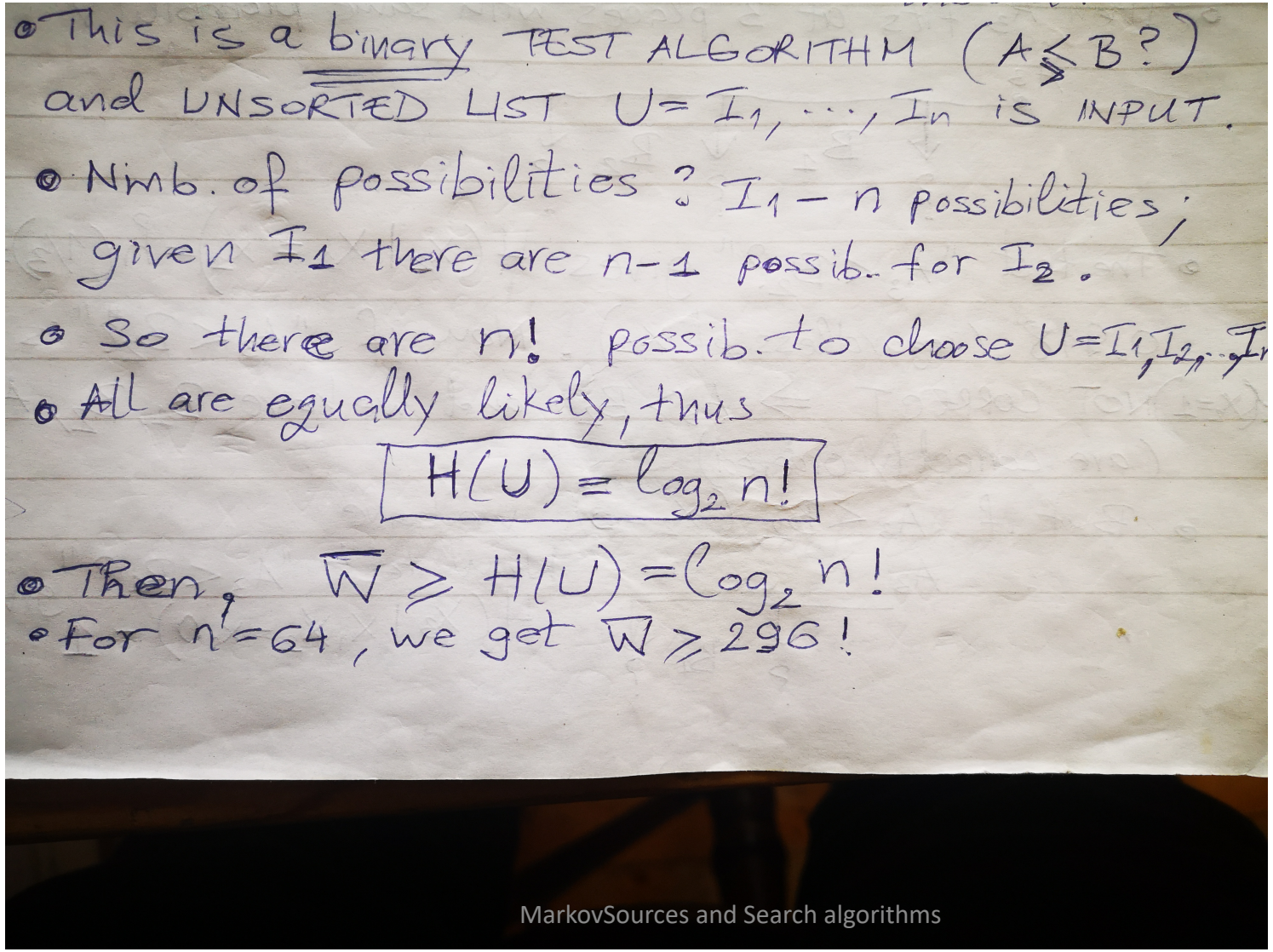
Hence

$$H_\infty(X) = \tfrac{15}{43}h(\tfrac{1}{3}) + \tfrac{15}{43}h(\tfrac{1}{4}) + \tfrac{12}{43}h(\tfrac{1}{2}) \approx 0.9013 \text{ bit/source symb}$$

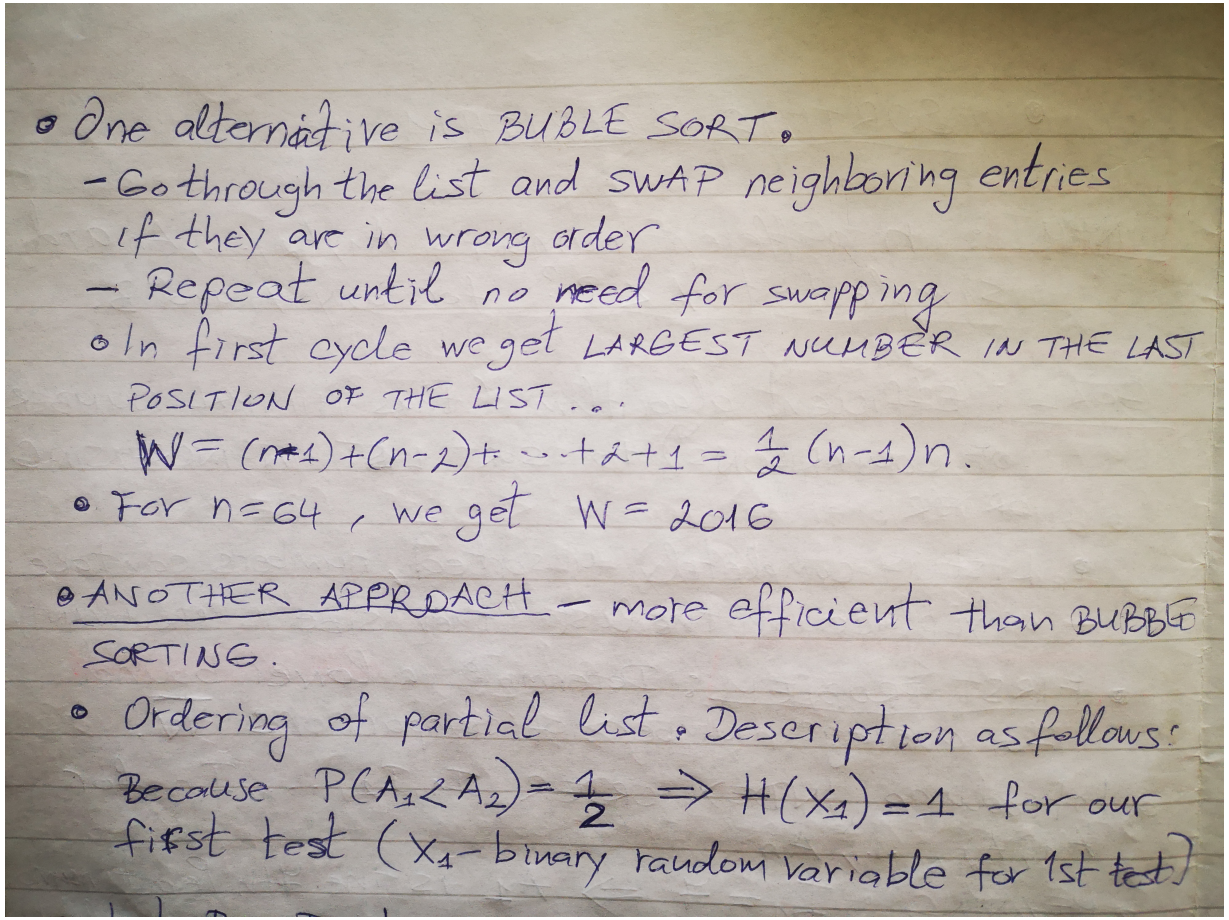
Sorting algorithms



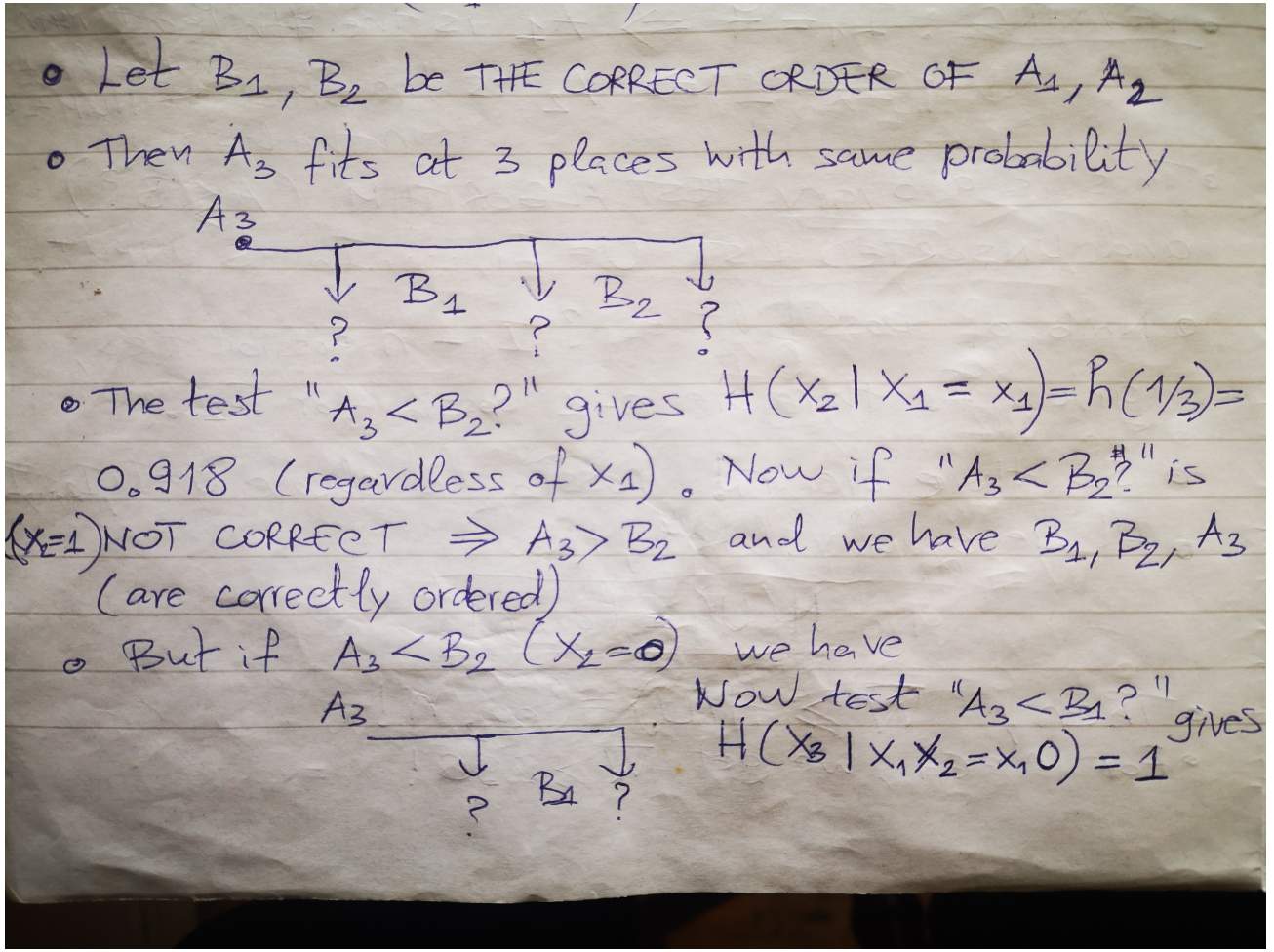
Sorting algorithms II



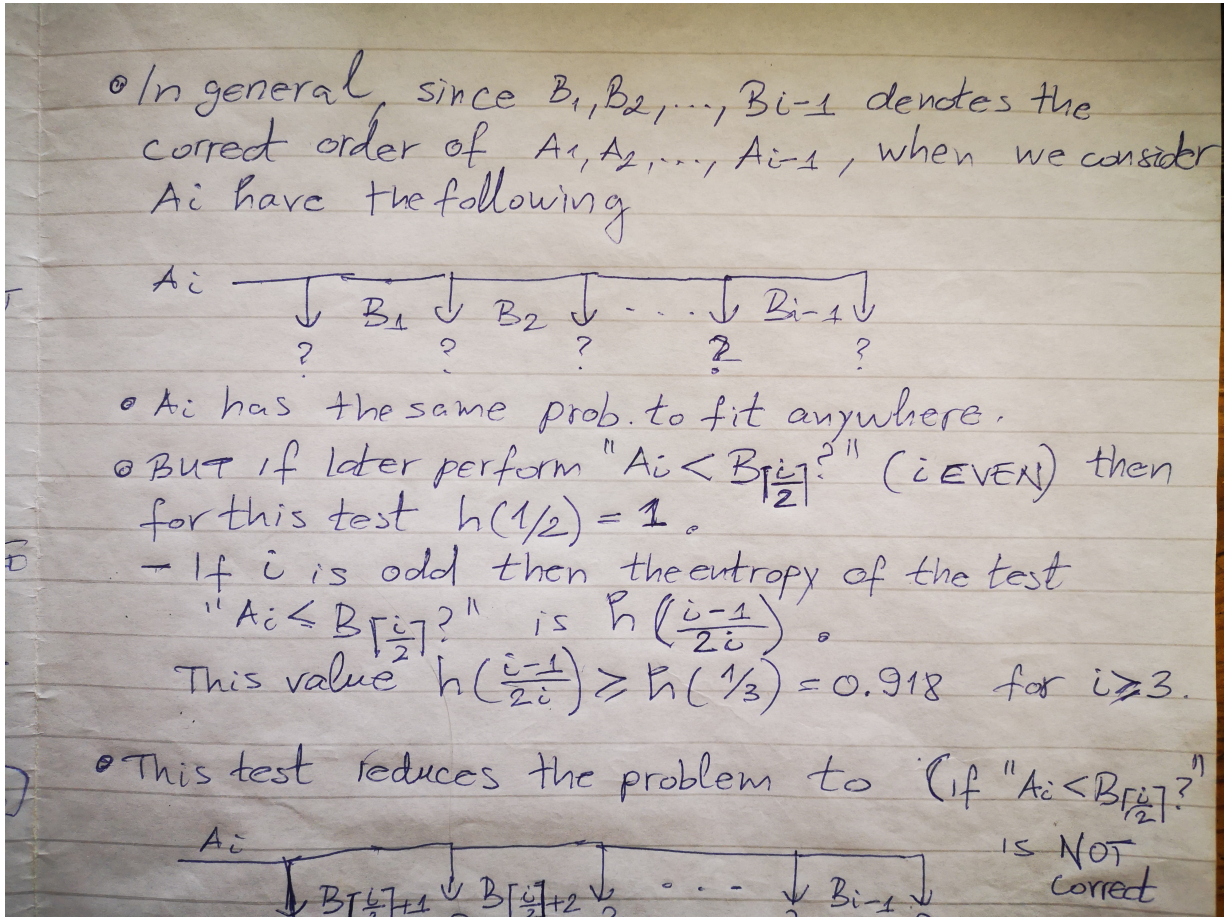
Buble sort algorithm



Description of the algorithm



General conclusions



Average number of tests - comparison