Fundamentals Of Information Science	**2022 Spring**

Homework4

学生：华园（202000120027）） 时间：*2022.3.17*

Problem 1.

Run-length encoding Run-length encoding is a popular variable-length lossless compressor used in fax machines, image compression, etc. Consider compression of S^n- an i.i.d. binary source with very small 1/64 of being 1 using run-length encoding f: A chunk of consecutive r <=127 zeros (resp. ones) is encoded into a zero (resp. one) followed by an 7-bit binary encoding of r (If there are > 127 consecutive zeros then two or more 8-bit blocks will be output). Compute the average achieved compression rate

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[l\left(f\left(S^n\right)\right)\right]$$

How does it compare with the optimal lossless compressor?

Solution.

S^n	$f(S^n)$	P	Length	normalize
0	0000 0001	$\frac{63}{64}*\frac{1}{64}$	8	8
1	1000 0001	$\frac{1}{64}*\frac{63}{64}$	8	8
00	0000 0010	$(\frac{63}{64})^2*\frac{1}{64}$	8	4
11	1000 0010	$(\frac{1}{64})^2*\frac{63}{64}$	8	4
...	...	...	...	...
00000..(127)	0111 1111	$(\frac{63}{64})^{127}$	8	8/127
11111..(127)	1111 1111	$(\frac{1}{64})^{127}$	8	8/127
000000...(k)	....	$(\frac{63}{64})^k$	$(\lfloor\frac{k}{127}+1\rfloor)*8$	...
111111...(k)	....	$(\frac{1}{64})^k$	$(\lfloor\frac{k}{127}+1\rfloor)*8$	...

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}E\left[l\left(f\left(S^n\right)\right)\right] &= \lim_{n\to\infty}\frac{1}{n}\left(\frac{63}{64^2}\cdot 8n+\frac{1}{64^2}\cdot 8n+\frac{63^2}{64^3}+4n+\frac{63}{64^3}\cdot 4n+\ldots+\left(\frac{63}{64}\right)^{127}\cdot\frac{8n}{127}+\left(\frac{1}{64}\right)^{127}\cdot\frac{8n}{127}\right)\\ &=\sum\nolimits_{k=1}^{126}\left[\frac{1}{64}\cdot\left(\frac{63}{64}\right)^k\cdot\frac{8}{k}+\frac{63}{64}\cdot\left(\frac{1}{64}\right)^k\cdot\frac{8}{k}\right]+\left(\frac{63}{64}\right)^{127}\cdot\frac{8}{127}+\left(\frac{1}{64}\right)^{127}\cdot\frac{8}{127}\\ &\approx 0.646\end{aligned}\tag{1.1}$$

program:

```
a=0
for i in range(1,127):
m=(1/64)*(63/64)**i*(8/i)+(63/64)*(1/64)**i*(8/i)
a=a+m
a=a+(63/64)**127*(8/127)+(1/64)**127*(8/127)
print(a)#0.6462243257022445
```

because δ_n is an i.i.d. binary source,so the entropy:

$$
\begin{aligned}
H(X) &= -n\sum_{n=1}^{\infty} p(x)log_2 p(x) \\
&= -n(\frac{1}{64}log_2\frac{1}{64} + \frac{63}{64}log_2\frac{63}{64}) \\
&= 0.116n
\end{aligned}
$$

$$
\begin{aligned}
if\ 0.116n < L = 0.646n < 0.116n + 1 \\
so\ n < 2
\end{aligned}
$$

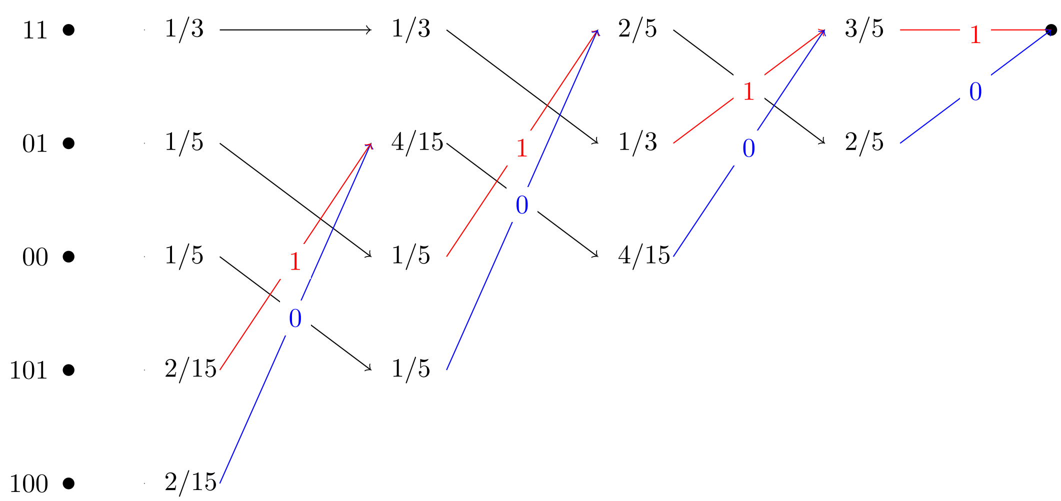
Therefore, we can know that the overall performance of run length coding is worse than the optimal code.

Problem 2.

Find the binary Huffman code for the source with probabilities (1/3,1/5,1/5,2/15,2/15). Argue that this code is also optimal for the source with probabilities (1/5 ,1/5,1/5, 1/5,1/5).

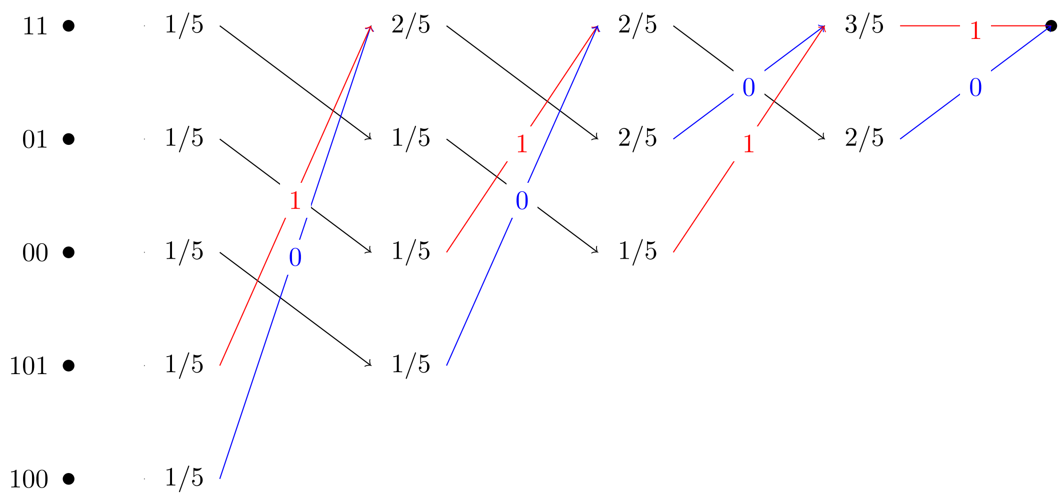
Solution.

for probabilities (1/3,1/5,1/5,2/15,2/15).



C(1/3)=11.C(1/5)=01,C(1/5)=00,C(2/15)=101,C(2/15)=100

for probabilities (1/5 ,1/5,1/5, 1/5,1/5).



We can proof that this code is also optimal for the source with probabilities (1/5 ,1/5,1/5, 1/5,1/5).

Problem 3.

Codes. Which of the following codes are
(a) Uniquely decodable?
(b) Prefix codes?

$$C1 = \{00, 01, 0\}$$
$$C2 = \{00, 01, 100, 101, 11\}$$
$$C3 = \{0, 10, 110, 1110, ...\}$$
$$C4 = \{0, 00, 000, 0000\}$$

Solution.

(a)C2 and C3 are uniquely decodable.
(For C1, 00 can be composed of 0, so it is not uniquely decodable.For C4, it is obvious that it is not uniquely decodable.)

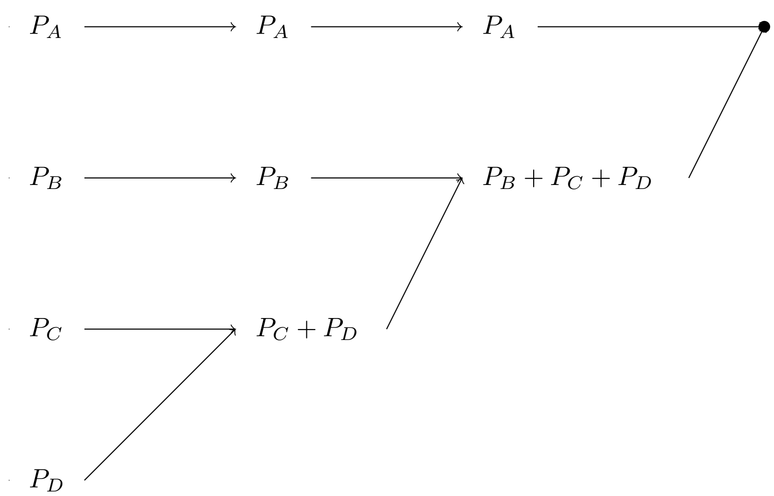
(b) C2 and C3 are prefix codes.
(All codewords in C2 and C3 are not prefixes of other codewords, so they are prefix codes.)

Problem 4.

Huffman is given four symbols, A, B, C, and D. The probability of symbol.A occurring is pA, symbol B is pB, symbol C is pC and symbol D is pD, with pA >= pB >= pC >= pD. Write down a single condition (equation or inequality) that is both necessary and sufficient to guarantee that, A will be encoded using exactly one bit. Explain your answer.

Solution.

If A would be encoded using exactly one bit,we can get:



From this, we can obtain that the constraint condition is:

$$P_A \geqslant P_C + P_D.$$