1) Problem

A coin is flipped until the first tail occurs. Let X denote the number of flips required. What is $H(X)$?

Solution

$$p(X=k)=0.5^{k-1}0.5=0.5^k$$

$$H(X)=\sum_{k=1}^{\infty}0.5^k\, ld\left(\frac{1}{0.5^k}\right)=\sum_{k=1}^{\infty}k0.5^k=2$$

2) Problem

Given a four-symbol random source with the following joint distribution of two consecutive symbols X and Y, as follows:

	X_1	X_2	X_3	X_4
Y_1	1/8	1/16	1/32	1/32
Y_2	1/16	1/8	1/32	1/32
Y_3	1/16	1/16	1/16	1/16
Y_4	1/4	0	0	0

Calculate:

a.) H(X) and H(Y)
b.) H($X|Y$) and H($Y|X$)
c.) H(X,Y)
d.) I(X,Y)

Solution

The marginal distributions are X (1/2,1/4,1/8,1/8) and Y (1/4,1/4,1/4,1/4)
The conditional distributions:

$$p(y|x)=\frac{p(x,y)}{p(x)},\quad p(x|y)=\frac{p(x,y)}{p(y)}$$

$$P(Y\mid X)=\begin{pmatrix}1/4 & 1/4 & 1/4 & 1/4\\ 1/8 & 1/2 & 1/4 & 1/4\\ 1/8 & 1/4 & 1/2 & 1/2\\ 1/2 & 0 & 0 & 0\end{pmatrix}$$

$$P(X\mid Y)=\begin{pmatrix}1/2 & 1/4 & 1/8 & 1/8\\ 1/4 & 1/2 & 1/8 & 1/8\\ 1/4 & 1/4 & 1/4 & 1/4\\ 1 & 0 & 0 & 0\end{pmatrix}$$

a)

$$\mathrm{H}(X)=\sum_x p(x)ld\left(\frac{1}{p(x)}\right)=\frac{1}{2}ld(2)+\frac{1}{4}ld(4)+\frac{1}{8}ld(8)+\frac{1}{8}ld(8)=\frac{1}{2}+\frac{2}{4}+\frac{3}{8}+\frac{3}{8}=\frac{7}{4},$$

H(*Y*)=2

b)

$$H(Y|X)=\sum_x p(x)\sum_y p(y|x)ld\left(\frac{1}{p(y|x)}\right)=\sum_x\sum_y p(x,y)ld\left(\frac{1}{p(y|x)}\right)=$$

$$=\frac{1}{8}ld\left(\frac{1}{1/4}\right)+\frac{1}{16}ld\left(\frac{1}{1/8}\right)+\frac{1}{16}ld\left(\frac{1}{1/8}\right)+\frac{1}{4}ld\left(\frac{1}{1/2}\right)+$$

$$+\frac{1}{16}ld\left(\frac{1}{1/4}\right)+\frac{1}{8}ld\left(\frac{1}{1/2}\right)+\frac{1}{16}ld\left(\frac{1}{1/4}\right)+$$

$$+\frac{1}{32}ld\left(\frac{1}{1/4}\right)+\frac{1}{32}ld\left(\frac{1}{1/4}\right)+\frac{1}{16}ld\left(\frac{1}{1/2}\right)+$$

$$+\frac{1}{32}ld\left(\frac{1}{1/4}\right)+\frac{1}{32}ld\left(\frac{1}{1/4}\right)+\frac{1}{16}ld\left(\frac{1}{1/2}\right)=$$

$$=\frac{2}{8}+\frac{3}{16}+\frac{3}{16}+\frac{1}{4}+\quad\frac{2}{16}+\frac{1}{8}+\frac{2}{16}+\quad\frac{2}{32}+\frac{2}{32}+\frac{1}{16}+\quad\frac{2}{32}+\frac{2}{32}+\frac{1}{16}$$

$$=\frac{13}{8}$$

and similarly $H(X|Y)=\sum_x\sum_y P(X=x,Y=y)\cdot ld\left(1/P(X=x|Y=y)\right)=11/8$.

c) $H(X,Y)=27/8$

d) $I(X,Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)=7/4-11/8=3/8$

3) Problem

Given a memoryless source with the following distribution

$$p_1=0.7;\ \ p_2=0.2;\ \ p_3=0.1$$

a.) What is the theoretical limit of compression !

b.) Compress the source by Huffman coding: give the code and the average codelength !

c.) What is the average code-length if we compress the source by Shannon-Fano coding?

Solution

a) Information of a symbol is defined as:

$$I(x)=\log_2\left(\frac{1}{p(x)}\right)=ld\left(\frac{1}{p(x)}\right)=-ld\left(p(x)\right)$$

the entropy of a source is defined as:

$$H(X)=\mathrm{E}_X\left(I(x)\right)=\sum_x p(x)ld\left(\frac{1}{p(x)}\right)=-\sum_x p(x)ld\left(p(x)\right)$$

which is the "average information"

The theoretical limit of the compression is the entropy.

$$I(x_1) = ld\left(\frac{1}{p_1}\right) = -ld(p_1) = -ld(0.7) \cong 0.5147$$

$$I(x_2) = ld\left(\frac{1}{p_2}\right) = -ld(p_2) \cong 2.32$$

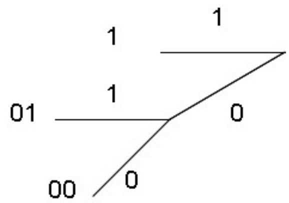
$$I(x_3) = ld\left(\frac{1}{p_3}\right) = -ld(p_3) \cong 3.321$$

$$H(X) = p_1 ld\left(\frac{1}{p_1}\right) + p_2 ld\left(\frac{1}{p_2}\right) + p_3 ld\left(\frac{1}{p_3}\right) = 0.3598 + 0.464 + 0.3321 = 1.1559$$

b) The Huffmann coding of the source:

Arrange the symbol probabilities in a decreasing order,
and add the two smallest term to create a new level of a binary tree.
Repeat it until reaching the root node.
The leaf nodes will be the associated codewords to the symbols with length equal to the level of node from the root in the tree.

0.7		1
0.2	0.3	
0.1		



The average code length is: $L = \mathrm{E}_x\left(l(x)\right) = \sum_x p(x) l(x)$

$$L_H = 1 \cdot 0.7 + 2 \cdot 0.2 + 2 \cdot 0.1 = 1.3$$

c) Using Shannon-Fano coding:

Generate $l(x) = \left\lceil ld\left(\frac{1}{p(x)}\right)\right\rceil$ for every x

Then populate a binary tree such that codewords can only be leaf nodes in the tree, to ensure prefixness. Choose the nodes with the appropriate depth in the tree corresponding to the length of the codeword computed before.

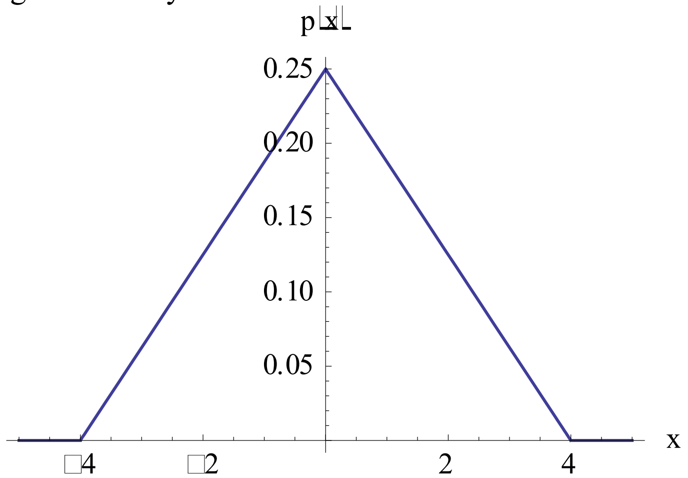
The average code length can be given without defining the actual codewords:

$$L_{SF} = l_1 p_1 + l_2 p_2 + l_3 p_3 + l_4 p_4 = \left\lceil ld\left(\frac{1}{p_1}\right)\right\rceil p_1 + \left\lceil ld\left(\frac{1}{p_2}\right)\right\rceil p_2 + \left\lceil ld\left(\frac{1}{p_3}\right)\right\rceil p_3 =$$

$$= 1 \cdot 0.71 + 3 \cdot 0.2 + 4 \cdot 0.1 = 1.7$$

4) Problem

After sampling a random signal which is stationary in the strong sense, the samples have the following the density function:



A four-level uniform (equidistant) quantization is applied on each sample in the interval $[-3,3]$.

a.) Describe the IT source model (the source alphabet and the corresponding probabilities)
b.) What is theoretical lower bound on compression.
c.) Give the average code-length in the case of Shannon-Fano coding!
d.) Develop Huffman coding!
e.) Compare the required dataspeed for transmitting the source information at sample frequency f_s=100MHz in case of Shannon-Fano and Huffman coding!

Solution

a) Due to uniform quantization the quantization levels are: $X = \{-3,-1,1,3\}$

$$p_1 = \int_{-\infty}^{-2} p(x)dx;\; p_2 = \int_{-2}^{0} p(x)dx;\; p_3 = \int_{0}^{2} p(x)dx;\; p_4 = \int_{2}^{\infty} p(x)dx$$

$$p_1 = \int_{-4}^{-2} \frac{x}{16} + 1/4dx = 1/8;\; p_2 = \int_{-2}^{0} \frac{x}{16} + 1/4dx = 3/8;$$

$$p_3 = \int_{0}^{2} -\frac{x}{16} + 1/4dx = 3/8;\; p_4 = \int_{2}^{4} -\frac{x}{16} + 1/4dx = 1/8;$$

b) $ld\left(\frac{1}{p_1}\right) = 3; ld\left(\frac{1}{p_2}\right) = 1{,}415; ld\left(\frac{1}{p_3}\right) = 1{,}415; ld\left(\frac{1}{p_4}\right) = 3$

$$H(X) = p_1 ld\left(\frac{1}{p_1}\right) + p_2 ld\left(\frac{1}{p_2}\right) + p_3 ld\left(\frac{1}{p_3}\right) + p_4 ld\left(\frac{1}{p_4}\right) =$$

$$= \frac{1}{8}\cdot 3 + \frac{3}{8}\cdot 1.415 + \frac{3}{8}\cdot 1.415 + \frac{1}{8}\cdot 3 = 1{,}81125$$

c) The average code length using Shannon-Fano coding is

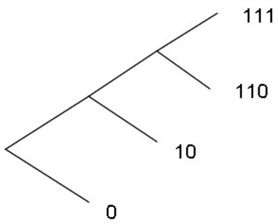
$$L_{SF} = l_1 p_1 + l_2 p_2 + l_3 p_3 + l_4 p_4 =$$

$$= \left\lceil ld\left(\frac{1}{p_1}\right)\right\rceil p_1 + \left\lceil ld\left(\frac{1}{p_2}\right)\right\rceil p_2 + \left\lceil ld\left(\frac{1}{p_3}\right)\right\rceil p_3 + \left\lceil ld\left(\frac{1}{p_4}\right)\right\rceil p_4 =$$

$$= 3\cdot\frac{1}{8} + 2\cdot\frac{3}{8} + 2\cdot\frac{3}{8} + 3\cdot\frac{1}{8} = \frac{18}{8} = 2.25$$

d)

0.375	0.375	**0.375**
0.375	**0.375**	0.615
0.125	0.25	
0.125		



$$L_H = 1\cdot\frac{3}{8} + 2\cdot\frac{3}{8} + 3\cdot\frac{1}{8} + 3\cdot\frac{1}{8} = \frac{15}{8} = 1.875$$

e) $Dataspeed^{SF} = 2.25\cdot 10^8 = 225 Mbps$; $Dataspeed^{Huff} = 1.875\cdot 10^8 = 187.5 Mbps$