Edgar
Kaziakhmedov

Information Theory
19.02.2019

Assignment 1

Problem 1

Let us suppose that we have the following possible events:

$$\{1, 01, 001, 0001, 00001, ..., \underbrace{00..001}_{k\,tosses}, \underbrace{00..001}_{k+1\,tosses}, ...\}$$

where each entry represents the outcomes of tossing a coin (1 - represents head, 0 is for tail). Assuming that probability of head occurring is equal to p and tail is p as well we can write down probabilities for each possible outcome corresponding to each event above:

$$\{p, p^2, p^3, p^4, p^5, ..., p^k, p^{k+1}, ...\}$$

So the entropy will be:

$$H(x) = -\sum_{i=1}^{\infty} p^i \log p^i = -p \log p \sum_{i=1}^{\infty} ip^{i-1} = -p \log p \sum_{i=1}^{\infty} \Big(p^i\Big)' = -p \log p \Big(\sum_{i=1}^{\infty} p^i\Big)' = -p \log p \Big(\frac{p}{p-1}\Big)'$$

Finally:

$$H(x) = -\frac{p \log p}{(1-p)^2}$$

Since $p = \frac{1}{2}$:

$$\boxed{H(x) = 2}$$

Problem 2

a) The more equiprobable samples the more entropy is, so under that logic if we have a vector with n probability components the entropy of it reaches maximum once the probability of each component is $\frac{1}{n}$ such that they are getting indistinguishable.
So the uniform distribution is **optimal** and:

$$H_{max} = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \boxed{\log n}$$

b) The same way of thinking is applicable here as well except we fix the first component to α. So for optimal distribution all the rest $n-1$ components should be uniformly distributed:

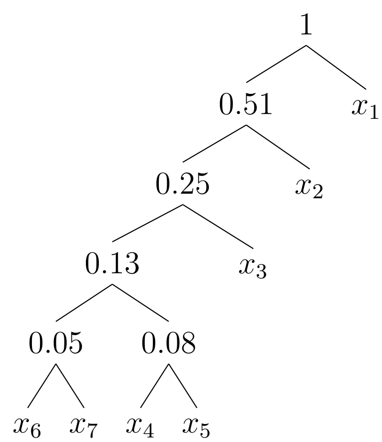
$$P = \Big\{p_1 = \alpha, p_i = \frac{1-\alpha}{n-1}, i = 2..n\Big\}$$

Thereby the maximal entropy will be the following:

$$H_{max} = -\alpha \log \alpha - \sum_{i=2}^{n} \frac{1-\alpha}{n-1} \log \frac{1-\alpha}{n-1} = \boxed{-\alpha \log \alpha - (1-\alpha) \log \frac{1-\alpha}{n-1}}$$

Problem 3

a) To construct binary Huffman tree we simple put all letters to the priority queue and extract two elements with the smallest probability. After taking two leafs we create simple tree where the parent node has a probability equal to a sum of probabilities of its leafs. Meanwhile, we mark each left branch with 0 and right branch with 1. After applying this algorithm we will get the following tree:



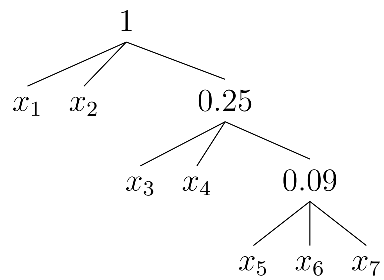
Now we can build up encoding table for each letter:

x_1	0
x_2	10
x_3	110
x_4	11100
x_5	11101
x_6	11111
x_7	11110

Finally the expected codelength is the following:

$$L_E = \sum_{i=1}^{7} w_i l_i = 0.49 + 0.52 + 0.36 + 0.2 + 0.2 + 0.15 + 0.1 = \boxed{2.02}$$

b) In the case of ternary Huffman code the situation is exactly the same except we take three elements from the queue and then do all the same set of operations, so we get the following Huffman tree:



Similarly build up encoding table for each letter:

x_1	0
x_2	1
x_3	20
x_4	21
x_5	220
x_6	221
x_7	222

So the expected codelengh:

$$L_E = \sum_{i=1}^{7} w_i l_i = 0.49 + 0.26 + 0.24 + 0.08 + 0.12 + 0.09 + 0.06 = \boxed{1.34}$$

Problem 4

a) According to the definition:

$$C_{max} = \max_{p(x)} I(X, Y)$$

where:

$$I(X, Y) = H(Y) - H(Y|X) = H(Y) - \sum_i P(x = i) H(Y|x = i)$$

Suppose that $\Pr(x = 0) = p$ and $\Pr(x = 1) = 1 - p$, then:

$$H(Y) = -\Pr(y = 0) \log \Pr(y = 0) - \Pr(y = 1) \log \Pr(y = 1)$$

Using law of total probability:

$$\Pr(y = 0) = \Pr(x = 1) \Pr(y = 0|x = 1) + \Pr(x = 0) \Pr(y = 0|x = 0) = p(\delta - 1) + 1$$

$$\Pr(y = 1) = \Pr(x = 1) \Pr(y = 1|x = 1) + \Pr(x = 0) \Pr(y = 1|x = 0) = p(1 - \delta)$$

Putting all together we will get:

$$H(Y) = -(p(\delta - 1) + 1)\log(p(\delta - 1) + 1) - p(1 - \delta)\log(p(1 - \delta))$$

Now let's write down the second term:

$$H(Y|X) = \Pr(x = 0)H(Y|x = 0) + \Pr(x = 1)H(Y|x = 1) =$$

$$(1-p)\Big(-\Pr(y = 0|x = 0)\log(\Pr(y = 0|x = 0)) - \Pr(y = 1|x = 0)\log(\Pr(y = 1|x = 0))\Big) +$$

$$+p\Big(-\Pr(y = 0|x = 1)\log(\Pr(y = 0|x = 1)) - \Pr(y = 1|x = 1)\log(\Pr(y = 1|x = 1))\Big) =$$

$$= p\Big(-\delta\log\delta - (1 - \delta)\log(1 - \delta)\Big) = pf(\delta)$$

And finally formula for mutual information:

$$I(X, Y) = -(p(\delta - 1) + 1)\log(p(\delta - 1) + 1) - p(1 - \delta)\log p(1 - \delta) + pf(\delta)$$

Now we will derivative of it and work it out:

$$\frac{dI}{dp} = -(\delta - 1)\log(p(\delta - 1) + 1) - (1 - \delta)\log p(1 - \delta) + f(\delta) = 0$$

After a set of optimization we obtain:

$$\boxed{p = \frac{1}{(1 - \delta)\Big(2^{\frac{-f(\delta)}{1-\delta}} + 1\Big)}}$$

Taking into account properties of mutual information the capacity of the channel is:

$$\boxed{C_{max} = I(p)}$$

b) Now things are getting more interesting as we take the limit of p with $\delta \to \infty$:

$$\lim_{\delta\to\infty} p(\delta) = \lim_{\delta\to\infty} \frac{1}{(1 - \delta)\Big(2^{\frac{-f(\delta)}{1-\delta}} + 1\Big)}$$

Let's replace the varible $1 - \delta$ with τ:

$$\lim_{\tau\to 0} \frac{1}{\tau\Big(2^{\frac{-f(1-\tau)}{\tau}} + 1\Big)}$$

Let's expand the function $f(\delta)$:

$$\lim_{\tau\to 0} \frac{1}{2^{\frac{-\log(1-\tau)}{\tau} + \log(1-\tau)} + 2^{\log\tau}}$$

Replacing logarithms with Taylor series reporesentaion gives us:

$$\lim_{\tau \to 0} \frac{1}{2^{-\frac{1}{\tau}\left(-\tau - \frac{\tau^2}{2} + \ldots\right) + \left(-\tau - \frac{\tau^2}{2}\right)} + 2^{\log \tau}}$$

So:

$$\boxed{\lim_{\tau \to 0} p(\delta) = \frac{1}{2}}$$

And:

$$\boxed{C\Big(p = \frac{1}{2}\Big) = 0}$$

Problem 5

First of all, let's write down matrix for the channel $(X \to Z)$:

$$\begin{pmatrix} \Pr(z=0|x=0) & \Pr(z=?|x=0) & \Pr(z=1|x=0) \\ \Pr(z=0|x=1) & \Pr(z=?|x=1) & \Pr(z=1|x=1) \end{pmatrix}$$

Or:

$$\begin{pmatrix} (1-p)(1-p_e) & p_e & p(1-p_e) \\ p(1-p_e) & p_e & (1-p)(1-p_e) \end{pmatrix}$$

According to the definition:

$$C_{max} = \max_{p(x)} I(X,Y)$$

Where:

$$I(X,Y) = H(Z) - H(Z|X)$$

Also we assume that:

$$P(x=1) = \pi$$

$$P(x=0) = 1 - \pi$$

Thereby:

$$H(Z) = -\Big(\pi p(1-p_e) + (1-\pi) + (1-\pi)(1-p)(1-p_e)\Big) \log \Big(\pi p(1-p_e) + (1-\pi) + (1-\pi)(1-p)(1-p_e)\Big) -$$

$$-p_e \log p_e -$$

$$-\Big(\pi(1-p)(1-p_e) + (1-\pi)p(1-p_e)\Big) \log \Big(\pi(1-p)(1-p_e) + (1-\pi)p(1-p_e)\Big)$$

$$H(Z|x=0) = -(1-p)(1-p_e)\log(1-p)(1-p_e) - p_e \log p_e - p(1-p_e)\log p(1-p_e)$$

$$H(Z|x=1) = -(1-p)(1-p_e)\log(1-p)(1-p_e) - p_e \log p_e - p(1-p_e)\log p(1-p_e)$$

Hence:

$$H(Z|X) = H(Z|x = 0)$$

As we need to maximize the entropy over X distribution we consider π to be equal to $\frac{1}{2}$ and derive the final answer:

$$\begin{aligned} C &= -(1-p_e)\log(1-p_e) + (1-p)(1-p_e)\log(1-p)(1-p_e) + p(1-p_e)\log p(1-p_e) = \\ &= -(1-p_e)\log(1-p_e) + (1-p_e) + (1-p)(p_e)\log(1-p_e) + p(1-p_e)\log(1-p_e) + \\ &\quad + (1-p)(1-p_e)\log(1-p) + p(1-p_e)\log p = \\ &= \boxed{\Big(1-p_e\Big)\Big((1-p)\log(1-p) + p\log p\Big)} \end{aligned}$$

Problem 6

The channel capacity of parallel Gaussian channel:

$$= \frac{1}{2}\log(1 + \frac{P_i}{N_i})$$

Also we need to maximize:

$$\sum_{i=1}^{k} \log(1 + \frac{P_i}{N_i})$$

over P_i such that (P=11):

$$\sum_{i=1}^{k} P_i = P,\ P_i \geq 0$$

To find P_i we can use Lagrange method:

$$L = \frac{1}{2}\sum_{i=1}^{6} \log(1 + \frac{P_i}{N_i}) + \lambda(\sum_i P_i - 11)$$

$$\frac{dL}{dP_i} = 0 \implies \frac{dL}{dP_i} = \frac{1}{2}\frac{1}{N_i + P_i} + \lambda = 0 \implies -2\lambda = \frac{1}{N_i + P_i}$$

$$P_i = -N_i - \frac{1}{2\lambda},\ \sum_{i=1}^{6} P_i = -\frac{3}{\lambda} - 11 = 11 \implies \lambda = -\frac{3}{22}$$

Then:

$$P_1 = P_2 = P_3 = \frac{8}{3},\ P_4 = \frac{5}{3},\ P_5 = P_6 = \frac{2}{3}$$

So the capacity:

$$C = \frac{1}{2}(3\log(1 + \frac{8}{3}) + \log(1 + \frac{5}{6}) + 2\log(1 + \frac{2}{9}) = \boxed{3.5384}$$

Problem 7