Information Theory: Assignment #2

Due on February 12, 2014 at 3:10pm

Marek Smieja

Szymon Maszke

Problem 1

Give one example of encoding from each group mentioned in lectures:

Solution

- **Non non-singular encoding** (coding function, each symbol can be encoded in any way [resulting codes can be equal]
  Example:

$$S = \{a, b, c, d\}, A = \{0\}, \phi : S \to A* \tag{1}$$
$$\phi(a) = 0 \tag{2}$$
$$\phi(b) = 0 \tag{3}$$
$$\phi(c) = 0 \tag{4}$$
$$\phi(d) = 0 \tag{5}$$

- **Non-singular encoding** (injective coding function [resulting codes for each symbol cannot be equal]), fulfilling the property:

$$s_1 \neq s_2 \implies \phi(s_1) \neq \phi(s_2) \tag{6}$$

  Example:

$$S = \{a, b, c, d\}, A = \{0, 1, 2\}, \phi : S \to A* \tag{7}$$
$$\phi(a) = 0 \tag{8}$$
$$\phi(b) = 1 \tag{9}$$
$$\phi(c) = 10 \tag{10}$$
$$\phi(d) = 2 \tag{11}$$

  Codes do not have to be uniquely decodable!

- **Fixed length codes** (same number of bits used for each encoding)
  Example (no compression, uniquely decodable):

$$S = \{a, b\}, A = \{0, 1\}, \phi : S \to A* \tag{12}$$
$$\phi(a) = 1 \tag{13}$$
$$\phi(b) = 0 \tag{14}$$

- **Uniquely decodable** (Code is uniquely decodable if there is only one series able to produce it)
  Example: above
  Definitions: Code is uniquely decodable if it's extensions is non-singular, namely:

$$\phi(s_1, s_2, \ldots, s_k) := \phi(s_1)\phi(s_2)\ldots\phi(s_k) \tag{15}$$

Problem 1 continued on next page. . .

- **Prefix codes** (None of the codes are the prefix of other codes)

$$S = \{a, b, c, d\}, A = \{0, 1\}, \phi : S \to A* \tag{16}$$
$$\phi(a) = 00 \tag{17}$$
$$\phi(b) = 10 \tag{18}$$
$$\phi(a) = 11 \tag{19}$$
$$\phi(b) = 01 \tag{20}$$

Problem 2

Assume m-element coding alphabet. We want uniquely decodable code with lengths $l = (1, 1, 2, 3, 2, 3)$. What is the minimum value of m we can choose? Find encoding for this length.

Solution

Part A

To resolve this problem we have to use **Kraft's inequality** given by following equation:

$$\sum_{i=1}^{r} m^{-li} \leq 1 \tag{21}$$

where:

- r count of elements in coding alphabet A
- l_i length of each element in coding alphabet A
- m size of coding alphabet (and sought variable)

Substituting our example to above equation gives us:

$$\frac{2}{m^2} + \frac{2}{m^3} + \frac{2}{m^4} \leq 1 \tag{22}$$
$$\tag{23}$$

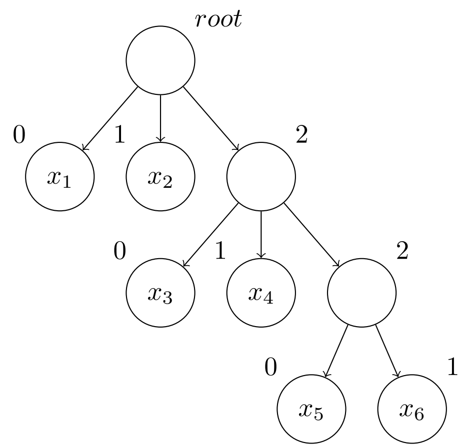
Solving above unequality for m gives us $\mathbf{m = 2.919}$. Size of the coding alphabet can only take integer numbers, hence:

$$m = \lceil 2.919 \rceil = 3 \tag{24}$$

Problem 2 continued on next page. . .

Part B

To create prefix code with minimal size of coding alphabet, we have to draw tree and apply codes from the left side (starting with the shortest ones)



Obtained codes from the tree are given by the following vector:

$$[\mathbf{0,\ 1,\ 20,\ 21,\ 220,\ 221}] \tag{25}$$

Problem 4

Check continuity and differentiability of the function:

$$sh(x) = \begin{cases} 0 & \text{if } x = 0 \\ -xlogx & \text{otherwise} \end{cases} \tag{26}$$

Solution

With the following function we have one point suspect od discontinuity, namely **0**, so we have to check limes from left and right side.

Part A

1. Transform $-xlogx$ to $-\frac{logx}{x^{-1}}$ so we can use l'Hospital rule
2. $$\lim_{x\to 0} -\frac{logx}{x^{-1}} \underset{\left[\frac{0}{0}\right]}{\overset{H}{=}} \lim_{x\to 0^+} \frac{x^{-1}}{x^{-2}} = \lim_{x\to 0^+} x^{-1} * x^2 = \lim_{x\to 0^+} x = 0 \tag{27}$$
3. $$\lim_{x\to 0} -\frac{logx}{x^{-1}} \underset{\left[\frac{0}{0}\right]}{\overset{H}{=}} \lim_{x\to 0^-} \frac{x^{-1}}{x^{-2}} = \lim_{x\to 0^-} x^{-1} * x^2 = \lim_{x\to 0^-} x = 0 \tag{28}$$
4. Left and right side limes is equal to zero and $sh(x)$ in zero is equal to zero, therefore the function is continuous and differentiable

Problem 5

Find maximum of the function $sh(x) + sh(1-x)$ with $x \in [0, 1]$ and draw it.

Solution

First of all we need the derivative of $sh(x) + sh(1-x)$, which is:

$$\frac{\mathrm{d}}{\mathrm{d}x}(sh(x)) + \frac{\mathrm{d}}{\mathrm{d}x}(sh(1-x)) = \begin{cases} 0 & \text{if } x = 0 \\ -xlogxf''(x) - (log(x)+1)f'(x) + log(1-x) + 1 & \text{otherwise} \end{cases} \tag{29}$$

Comparing derivative to zero gives us:

$$-xlogxf''(x) - (log(x)+1)f'(x) + log(1-x) + 1 = 0$$
$$x = \frac{1}{2} - \frac{\sqrt{e^2-4}}{2e}$$

And nobody in their right mind will calculate all of it by hand.

Problem 6

There is an event $X = x$ occuring with probability $p_x = P(X = x)$. Let $I(p(x)) = -\log_2 p(x)$ be the amount of information in event.
We play a game, where we choose one integer (x) number in the range $[0, 63]$. Opponent has to guess foremenetioned integer asking questions with only two possible answers (Yes or No).

- How many questions have to be asked to know the value of X? What's the information contained in the answer for each of those questions? What's the amount of information we get after asking all of the questions?
- Assume opponent is asking questions "Whether this number is y?", where y ia an integer we didn't ask for yet. We guessed the value of x after asking n questions. What is the information contained in each of those answers? Consider both positive and negative answers.
- Calculate the information received after n questions

Solution

Part A

We have to make following assumption to give meaningful answers:

Assumption: Choice of each integer in the range (0,63) is equally likely and follows uniform distribution (maximum entropy)

For easier comprehension I will introduce two easy notations.

- Maximum amount of information will be given for the smallest probability for random variable, e.g.
$$\lim_{p(i)\to 0^+} -\log_2 p(i) = \infty \tag{30}$$
- Smallest amount of information is given by:
$$\lim_{p(i)\to \infty} -\log_2 p(i) = 0 \tag{31}$$
- Information about X is equal to it's entropy, e.g.:
$$I(X; X) = H(X) \tag{32}$$

Part B

Solutions to the problems:

- Possible answers ('Yes' or 'No') follow Bernoulli distribution, hence maximum entropy is given by $p(y) = \frac{1}{2}$, where y is our answer. The only possibility for $p(y) = \frac{1}{2}$ is when both answers are Which gives us $x = 0$ and $y = 0$ equally likely. With set consisting of 64 numbers we can divide it into two parts with 32 elements each by asking question: 'Is the number bigger than a', where a is the center value of the set (in this case 31). To certainly receive the answer for set with n sorted values and asking binary questions (which allow to divide the set) we have to ask:
$$\log_2(n) \tag{33}$$
questions (**6 in this case**).
For those six questions we get a total information amount received equal to six as well:
$$6 * \log_2(\frac{1}{2}) = 6 \tag{34}$$
Which sums up to total count of questions we have to ask.

Problem 6 continued on next page. . .

- If opponent guesses the answer right away, the amount of information he receives is equal to:

$$-\log_2(\frac{1}{64}) = 6 \tag{35}$$

  With each answer 'No' amount of information received drops by $\frac{\log(\frac{n-1}{n})}{\log(2)}$, where n is the size of set for current guess, and $n-1$ the size of set for the guess before.

  Same situation applies to all negative guesses with sudden 'Yes' followed by it.

- For information contained in guess after n tries we can provide the generic formula:

$$-\log_2(\frac{1}{m-n}) \tag{36}$$

  where m is the count of set in the beginning, given the answers are binary and integer's probability ($p(x)$) follows the uniform distribution.

Problem 7

Calculate extrema of $z = x + y$ on $x^2 + y^2 = 1$

Solution

- Let's calculate partial derivatives of the function:

$$\frac{\partial z}{\partial x} = 1 \tag{37}$$

$$\frac{\partial z}{\partial y} = 1 \tag{38}$$

  which gives us statnioray point $P = (1, 1)$ where the extremum might be located.

- Second order derivatives would return 0 and so would *det* for the matrix W so this test is inconclusive

- In this scenario we can pose it as optimization problem with Lagrangian:

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y) \tag{39}$$

$$\mathcal{L}(x, y, \lambda) = x + y - \lambda(x^2 + y^2 - 1) = -\lambda x^2 - \lambda y^2 + x + y - \lambda \tag{40}$$

- Now we have to calculate partial derivatives with respect to each of the parameters:

$$\frac{\partial \mathcal{L}}{\partial x} = -2x\lambda + 1 \tag{41}$$

$$\frac{\partial \mathcal{L}}{\partial y} = -2y\lambda + 1 \tag{42}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -x^2 - y^2 - 1 \tag{43}$$

- Creating system of equations with each equal to zero gives us those values:

$$x = -\frac{i}{\sqrt{2}} \wedge y = -\frac{i}{\sqrt{2}} \tag{44}$$

$$\vee \tag{45}$$

$$x = \frac{i}{\sqrt{2}} \wedge y = \frac{i}{\sqrt{2}} \tag{46}$$