

Solutions to First Colloquium in Information Theory, April 2021

1. (15p) The inhabitants of an island tell the truth one third of the time. They lie with probability $2/3$. On an occasion, after one of them made a statement, you ask another ‘was that statement true?’ and he says ‘yes’. What is the probability that the statement was indeed true?

Define:

- event T - the first statement is true
- event F - the first statement is false
- event Y - the second statement is ‘yes’.

Then we have $P(T) = 1/3$, $P(F) = 2/3$, $P(Y|T) = 1/3$ and $P(N|T) = 2/3$; and using Bayes’ theorem

$$P(T|Y) = \frac{P(Y|T)P(T)}{P(Y)} = \frac{P(Y|T)P(T)}{P(Y|T)P(T) + P(Y|F)P(F)} = \frac{\frac{1}{3} \cdot \frac{1}{3}}{\frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{2}{3}} = \frac{1}{5}$$

So, the probability that the statement was indeed true is $1/5$.

2. (15p) Let X and Y be two independent integer-valued random variables. Let X be uniformly distributed over $\{1, 2, \dots, 8\}$, and let $P(Y = k) = 2^{-k}$, $k \in \{1, 2, 3, \dots\}$.

- (a) Find $H(X)$.
 (b) Find $H(Y)$. (You might find $\sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2}$ useful.)
 (c) Find $H(X + Y, X - Y)$.

- (a) For a uniform distribution, $H(X) = \log_2 8 = 3$.
 (b) For a geometric distribution,

$$H(Y) = \sum_{k=1}^{\infty} 2^{-k} \log_2 \frac{1}{2^{-k}} = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \log_2 2^k = \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k = \frac{1/2}{(1 - 1/2)^2} = 2$$

- (c) Since $(X, Y) \rightarrow (X + Y, X - Y)$ is a one to one transformation, and since X and Y are independent,

$$H(X + Y, X - Y) = H(X, Y) = H(X) + H(Y) = 3 + 2 = 5$$

3. (25p)

- (a) The source coding theorem shows that the optimal code for a random variable X has an expected length less than $H(X) + 1$. Give an example of a random variable for which the expected length of the optimal code is close to $H(X) + 1$, i.e., for any $\epsilon > 0$, construct a distribution for which the optimal code has expected length $L > H(X) + 1 - \epsilon$.

- (b) Consider a random variable Y which takes on four values with probabilities $(\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12})$.
- Construct a Huffman code for this random variable.
 - Show that there exist two different sets of optimal lengths for the codewords.
 - Are there optimal codes with codeword lengths for some symbols that exceed the information content of the symbol $\lceil \log_2 \frac{1}{p(y)} \rceil$?
- (a) There is a trivial example that requires almost 1 bit above its entropy. Let X be a binary random variable with probability of $X = 1$ close to 1. Then entropy of X is close to 0, but the length of its optimal code is 1 bit, which is almost 1 bit above its entropy.
- (b) i. Construct a Huffman tree. Examples of possible Huffman code for Y are $\{00, 01, 10, 11\}$ and $\{1, 01, 001, 000\}$.
- Both set of lengths 1,2,3,3 and 2,2,2,2 satisfy the Kraft inequality, and they both achieve the same expected length (2 bits) for the above distribution. Therefore they are both optimal.
 - The symbol with probability $1/4$ (from i.) has a Huffman code of length 3, which is greater than $\lceil \log_2 \frac{1}{p(y)} \rceil = \lceil \log_2 4 \rceil = 2$.

4. (25p)

- (a) Using Lempel–Ziv–Welch algorithm compress the following text: "bababcccbababaaaac" with the decimal initial dictionary

a	1
b	2
c	3

- (b) The source of information A generates the symbols $A = \{a, b, c\}$ with the probabilities $P_A = \{0.5, 0.3, 0.2\}$. Using arithmetic coding encode the message "bacb".

- (a) We form a table, keeping track of the input, output and dictionary.

Input	Output	Dictionary
b	2	ba - 4
a	1	ab - 5
ba	4	bab - 6
b	2	bc - 7
c	3	cc - 8
cc	8	ccb - 9
bab	6	baba - 10
ab	05	aba - 11
a	01	aa - 12
aa	12	aaa - 13
a	01	ac - 14
c	03	

So, writing the output column as a string, we get that the LZW encoding of bababcc-cbababaaaac is 21423860501120103.

- (b) First we find the lower and the upper end of the interval corresponding to bach.

$$l = 1 \cdot 0.5 + 0.3 \cdot 0 + 0.3 \cdot 0.5 \cdot (0.5 + 0.3) + 0.3 \cdot 0.5 \cdot 0.2 \cdot (0.5) = 0.635$$

So, the lower end of the interval is 0.635, and hence the upper end is $L = 0.635 + 0.3 \cdot 0.5 \cdot 0.2 \cdot 0.3 = 0.644$.

Now we need to find a binary representation of a number lying in the interval $[l, L)$. We do that in the following way.

$1/2 = 0.5 < 0.635$ add 1
 $0.5 + 1/4 = 0.75 > 0.644$ add 0
 $0.5 + 1/8 = 0.625 < 0.635$ add 1
 $0.625 + 1/2^4 = 0.6875 > 0.644$ add 0
 $0.625 + 1/2^5 = 0.65625 > 0.644$ add 0
 $0.625 + 1/2^6 = 0.640625$ in the interval - add 1 and stop.

So, and encoding of bach is 101001.

5. (20 p) In several proofs we have used the so-called relative distance (Kullback-Lebler distance) between two probability distributions $p(x)$ (true one) and $q(x)$ (estimated one). This distance was defined as $D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)}$.

- (a) Using Jensen's inequality which claims that $E[f(X)] \geq f(E[X])$ (valid for convex functions f) prove that $D(p||q) \geq 0$. Hint: Consider $-\log_2$ as function f .
- (b) Show that the mutual information $I(X;Y)$ can be expressed as $I(X;Y) = H(X) + H(Y) - H(X,Y)$ and therefore by definition of the (joint) entropy can be expressed as

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y) \log_2 \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}$$

- (c) Use the above parts to show that $I(X;Y) \geq 0$.

Check your lecture slides/notes for a solution of this problem.