Applied Mathematics Exercise 1

Erik Paskalev

September 10, 2024

1 Is there more information in a string of 10 letters from {A,...,Z} or in a string of 26 numbers from {0,...,9} ?

Answer: The information in the string of 10 letters is the sum of information of each letter so the total is: $10*(-log_2(\frac{1}{26})) \approx 47.004$ **bits. Analogously for the 26 numbers we sum the information of each number:** $26*(-log_2(\frac{1}{10})) \approx 86.370$ **bits. So we can conclude that the string of 26 numbers contains more information.**

.

2 Calculate the entropy of each of the following variables.

(a) Pixel values in an image whose possible grey values are all the integers from 0 to 255 with uniform probability.

Answer: The random variable ranges from 0 to 255 and the probability distribution is uniform so the entropy is:

$$-\sum_{i=0}^{255} \frac{1}{256} log_2(\frac{1}{256}) = -log_2(\frac{1}{256}) = 9 \text{ bits} \tag{1}$$

(b) A population of persons classified by whether they are older, or not older, than the population's median age.

Answer: The probability distribution consists of two outcomes - above median age and below each with 1/2 probability. Thus the entropy is:

$$-\sum_{i=1}^{2}\frac{1}{2}log_2(\frac{1}{2}) = -log_2(\frac{1}{2}) = 1 \text{ bit} \tag{2}$$

(c) Gender in a tri-sexual insect population, whose three genders occur with probabilities $\frac{1}{4}, \frac{1}{4}$ and $\frac{1}{2}$

Answer: The probability distribution consists of 3 outcomes with probabilities $\frac{1}{4}, \frac{1}{4}$ and $\frac{1}{2}$. Thus the entropy is:

$$-2 * \frac{1}{4}log_2(\frac{1}{4}) - \frac{1}{2}log_2(\frac{1}{2}) = \frac{3}{2} \text{ bits} \tag{3}$$

(d) Humans classified according to whether they are mammals or not.

Answer: Since all humans are mammals the probability distribution consists of one outcome with probability 1. Thus the entropy is:

$$-log_2(1) = 0 \text{ bits} \tag{4}$$

3 Let X denote the outcome of rolling a die.

(a) Determine the entropy of X.

Answer: Dice are six-sided and uniformly distributed. Thus the entropy is:

$$-\sum_{i=1}^{6}\frac{1}{6}log_2(\frac{1}{6}) = -log_2(\frac{1}{6}) = log_2(6) \approx 2.584 \text{ bits} \tag{5}$$

(b) Calculate the expected number of bits per outcome if you encode the outcomes with the Huffman coding.

Answer: The Huffman encoding of the die is 00 for 1, 01 for 2, 100 for 3, 101 for 4, 110 for 5, 111 for 6 Thus the entropy is:

$$2 * \frac{1}{6} * 2 + 4 * \frac{1}{6} * 3 \approx 2.666 \text{ expected bits} \tag{6}$$

4 There are two horse races taking place. In the first race 4 horses participate, of which two have probability $\frac{3}{8}$ on victory, while the other two win with probability $\frac{1}{8}$. In the second race 7 horses partake. One of them is clearly the favorite with a victory probability of $\frac{5}{8}$. All other horses have equal winning odds.

(a) Calculate the entropy for both races exactly, and simplify as much as possible. In which race the uncertainty is the largest?

Answer: The distribution for the first race is: $\frac{3}{8}, \frac{3}{8}, \frac{1}{8}, \frac{1}{8}$. and for the second: $\frac{5}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}$. Thus the entropy for both is:

$$-2 * \frac{3}{8} log_2(\frac{3}{8}) - 2 * \frac{1}{8} log_2(\frac{1}{8}) = \frac{3}{4} log_2(\frac{8}{3}) + \frac{1}{4} log_2(8) \approx 1.811 \text{ bits for race 1} \tag{7}$$

$$-\frac{5}{8} log_2(\frac{5}{8}) - 3 * \frac{1}{8} log_2(\frac{1}{8}) = \frac{5}{8} log_2(\frac{8}{5}) + \frac{3}{8} log_2(8) \approx 1.548 \text{ bits for race 2} \tag{8}$$

This makes race 1 more uncertain than race 2.

(b) Determine the Huffman code for both races.

Answer: The Huffman encoding of race 1 is 0 for horse 1, 10 for horse 2, 110 for horse 3, 111 for horse 4. The Huffman encoding of race 2 is 0 for horse 1, 10 for horse 2, 110 for horse 3, 111 for horse 4.

(c) Compute for both codes given in (b) the expected number of bits needed per outcome.

Answer: The expected number of bits for code 1:

$$\frac{3}{8} + 2 * \frac{3}{8} + 3 * \frac{1}{8} + 3 * \frac{1}{8} = 1.875 \text{ bits for race 1} \tag{9}$$

The expected number of bits for code 2:

$$\frac{5}{8} + 2 * \frac{1}{8} + 3 * \frac{1}{8} + 3 * \frac{1}{8} = 1.625 \text{ bits for race 2} \tag{10}$$

(d) Are the codes found in (b) optimal?

Answer: No, based on the Shannon theorem we can derive that the lowest expected number of bits needs to be equal to the entropy of the distribution to be optimal which for both races 1 and 2 is false.

5 The outcomes of two (fair) dice are described by the random variables X and Y.

(a) Give the joint probability distribution of X and Y (in other words, if you roll two dice X and Y at the same time, what are the possible outcomes and what is the probability of each outcome?).

Answer: Since the two die are independent, X and Y are independent random variables. Thus the joint probability distribution of X and Y is:

$$X = (1, 2, 3, 4, 5, 6), P(X) = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}) \tag{11}$$

$$Y = (1, 2, 3, 4, 5, 6), P(Y) = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}) \tag{12}$$

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) = \frac{1}{6} * \frac{1}{6} = \frac{1}{36} \tag{13}$$

P(X,Y)	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

Table 1: Joint probability distribution visualization.

(b) Calculate the exact entropy H(X,Y) of this distribution, and check that H(X,Y) = H(X) + H(Y).

Answer: The entropy is:

$$H(X,Y) = -\sum_{i=0}^{6}\sum_{j=0}^{6} A_i B_j log_2(A_i B_j) =$$
$$= -\sum_{i=0}^{36} \frac{1}{36} log_2(\frac{1}{36}) = log_2(36) \approx 5.169 \text{ bits} \tag{14}$$

$$H(X) = -\sum_{i=0}^{6} \frac{1}{6} log_2(\frac{1}{6}) = log_2(6) \approx 2.584 \text{ bits} \tag{15}$$

$$H(Y) = -\sum_{i=0}^{6} \frac{1}{6} log_2(\frac{1}{6}) = log_2(6) \approx 2.584 \text{ bits} \tag{16}$$

$$H(X) + H(Y) \approx 2.584 + 2.584 \approx 5.168 \approx H(X,Y) \tag{17}$$

(c) Denote the sum of both outcomes by V, i.e. $V = X + Y$. Compute H(V) in 3 decimals and show that H(V) < H(X,Y).

Answer: The entropy of V is:

$$H(V) = H(X + Y) = -2 * \frac{1}{36} log_2(\frac{1}{36}) - 2 * \frac{1}{18} log_2(\frac{1}{18}) -$$
$$- 2 * \frac{1}{12} log_2(\frac{1}{12}) - 2 * \frac{1}{9} log_2(\frac{1}{9}) - \tag{18}$$
$$- 2 * \frac{5}{36} log_2(\frac{5}{36}) - \frac{1}{6} log_2(\frac{1}{6}) \approx 3.274 \text{ bits} < H(X,Y) \approx 5.169$$

5

X+Y	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Table 2: Visualization of all outcomes of X+Y

V	2	3	4	5	6	7	8	9	10	11	12
P(V)	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Table 3: Visualization of the probability distribution of V.

(d) Denote the minimum of the outcomes by W, i.e. W = minX,Y. Compute H(W) in 3 decimals and show that H(W) < H(X).

Answer: The entropy of W is:

min X,Y	1	2	3	4	5	6
1	1	1	1	1	1	1
2	1	2	2	2	2	2
3	1	2	3	3	3	3
4	1	2	3	4	4	4
5	1	2	3	4	5	5
6	1	2	3	4	5	6

Table 4: Visualization of all outcomes of W

$$
\begin{aligned}
H(W) = &-\frac{1}{36}log_2(\frac{1}{36}) - \frac{1}{12}log_2(\frac{1}{12}) - \frac{5}{36}log_2(\frac{5}{36}) - \frac{7}{36}log_2(\frac{7}{36}) - \\
&- \frac{1}{4}log_2(\frac{1}{4}) - \frac{11}{36}log_2(\frac{11}{36}) \approx 2.319 \text{ bits} < H(X) \approx 2.584
\end{aligned}
\tag{19}
$$

X-Y	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Table 5: Visualization of all outcomes of Z

(e) Denote the absolute difference between the outcomes by Z, i.e. Z = |X-Y|. Compute H(Z) in 3 decimals and show that H(Z) < H(X).

Answer: The entropy of Z is:

$$H(V) = H(X+Y) = -\frac{1}{18}log_2(\frac{1}{18}) - \frac{1}{9}log_2(\frac{1}{9}) - \frac{1}{6}log_2(\frac{1}{6}) - \frac{2}{9}log_2(\frac{2}{9}) -$$
$$- \frac{5}{18}log_2(\frac{5}{18}) - \frac{1}{6}log_2(\frac{1}{6}) \approx 2.441 \text{ bits} < H(X) \approx 2.584$$

(20)

X,X	H	T
H	p^2	$p(1-p)$
T	$(1-p)p$	$(1-p)^2$

Table 6: Visualization of all outcomes of 2 coin filps

6 Let X be the outcome of a biased coin. Suppose that $P(X = H) = p$ and $P(X = T) = 1 - p$ for certain $p > \frac{1}{2}$. If we encode the outcomes with zeroes (for head) and ones (for tails), then we need 100 bits to encode 100 outcomes. There is however a more efficient code if we group the outcomes in pairs of two.

(a) Give the corresponding Huffman code. During the construction of the binary tree, you may assume that $1 - p < p^2$.

Answer: The probability distribution over a pair of coin flips is a joint probability distribution over X twice.

To generate the Huffman code for this distribution we start by connecting $(1-p)^2$ with $p(1-p)$ since we know $1 - p < p^2 \longrightarrow p^2 > p(1-p) > (1-p)^2$ after the combining we get $(1-p)^2 + p(1-p) = p^2 - 2p + 1 + p - p^2 = 1 - p$ comparing it to the remaining probabilities $p^2 > (1 - p) > (1-p)p$ next we combine the next smallest $(1-p) + (1-p)p = 1 - p + p - p^2 = 1 - p^2$ and lastly $p^2 + 1 - p^2 = 1$ so the final Huffman code we get is 0 for p^2, 10 for $(1-p)p$, 110 for $p(1-p)$ and 111 for $(1-p)^2$.

(b) Suppose that we already know that the expected number of bits to encode 100 outcomes is equal to 70 if we use this more efficient code. What can you conclude about the entropy of X?

Answer: The only definitive statement we can make about X is that its entropy is below 0.7 bits.

(c) Determine p (both exactly and in 3 decimals) and the entropy of X (in 3 decimals).

Answer: (Assuming we are meant to use the assumptions from (b)) the expected number of bits is 70 over 100 outcomes which means that for a single pair of outcomes, the expected bits are 1.4 so:

8

$$E(\#bits) = 1.4 = p^2 + 2 * (1-p)p + 3 * p(1-p) + 3 * (1-p)^2 =$$
$$= p^2 + 2p - 2p^2 + 3p - 3p^2 + 3p^2 - 6p + 3 = -p^2 - p + 3 = 1.4 \tag{21}$$

$$E(\#bits) = -p^2 - p + 1.6 = 0$$
$$D = 1 + 6.4 = 7.4$$
$$p_1 = \frac{1 - \sqrt{7.4}}{-2} \approx 0.860$$
$$p_2 = \frac{1 + \sqrt{7.4}}{-2} \approx -1.860 < 0$$
$$\implies p = \frac{1 - \sqrt{7.4}}{-2} \approx 0.860 \tag{22}$$

$$H(X) = -\frac{1 - \sqrt{7.4}}{-2} log_2(\frac{1 - \sqrt{7.4}}{-2}) -$$
$$-(1 - \frac{1 - \sqrt{7.4}}{-2}) log_2(1 - \frac{1 - \sqrt{7.4}}{-2}) = \tag{23}$$
$$= 0.1869 + 0.3969 = 0.583 \text{ bits}$$

7 Let X be a random variable with a binomial distribution with parameters n and p, i.e. the probability of the ith outcome is equal to:

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \tag{24}$$

Show that the entropy of X is given by:

$$H(X) = -n(p log_2(p) + (1 - p) log_2(1 - p)) \tag{25}$$

Answer: The Binomial distribution is defined as the joint probability distribution of n independent Bernoulli distributions and because the Bernoulli distributions are independent of one another then the associativity property of Entropy holds for joint probability distributions

$$Y = (p, (1 - p))$$
$$H(X) = \underbrace{H(Y) + H(Y) + ... + H(Y)}_{n \text{ times}} \tag{26}$$

But we already know the entropy of the Bernoulli distribution

$$H(Y) = -(p log_2(p) + (1 - p) log_2(1 - p)) \tag{27}$$

so the formula for the Binomial distribution would be:

$$H(X) = \underbrace{H(Y) + H(Y) + ... + H(Y)}_{n \text{ times}}$$

$$H(X) = n * H(Y)$$

$$H(X) = -n(p\log_2(p) + (1-p)\log_2(1-p))$$

(28)

8 Suppose X is a continuous random variable with an exponential probability distribution. Then X has a probability density function given by

$$f(x) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}} \text{ for } x \geq 0 \tag{29}$$

Show that the entropy of X is equal to

$$H(X) = \log_2(\lambda e) \tag{30}$$

Answer: The entropy is:

$$
\begin{aligned}
H(X) &= -\int \frac{1}{\lambda}e^{-\frac{x}{\lambda}}\log_2(\frac{1}{\lambda}e^{-\frac{x}{\lambda}})dx = \\
&= e^{-\frac{x}{\lambda}}\log_2(\frac{1}{\lambda}e^{-\frac{x}{\lambda}}) - \frac{1}{\lambda ln(2)}\int -e^{-\frac{x}{\lambda}}dx = \\
&= e^{-\frac{x}{\lambda}}\log_2(\frac{1}{\lambda}e^{-\frac{x}{\lambda}}) - \frac{\lambda e^{-\frac{x}{\lambda}}}{\lambda ln(2)} + C = \\
&= \frac{e^{-\frac{x}{\lambda}}(ln(\frac{e^{-\frac{x}{\lambda}}}{\lambda}) - 1)}{ln(2)} + C = \\
&= \frac{e^{-\frac{x}{\lambda}}(ln(\frac{e^{-\frac{x}{\lambda}}}{\lambda}) - ln(e))}{ln(2)} = \\
&= \frac{e^{-\frac{x}{\lambda}}ln(\frac{e^{-\frac{x}{\lambda}}}{\lambda e})}{ln(2)} = \\
&= e^{-\frac{x}{\lambda}}\log_2(\frac{e^{-\frac{x}{\lambda}}}{\lambda e}) = \\
&= e^{-\frac{x}{\lambda}}\log_2(e^{-\frac{x}{\lambda}}) - e^{-\frac{x}{\lambda}}\log_2(\lambda e) = \\
&= \underbrace{e^{-\frac{x}{\lambda}}\log_2(e^{-\frac{x}{\lambda}}) - (e^{-\frac{x}{\lambda}} + 1)\log_2(\lambda e)}_{\text{somehow supposed to be 0}} + \log_2(\lambda e) = \text{ idk}
\end{aligned}
$$

(31)