Information and Communication Theory

2022

Problem Set 1

Department of Electrical and Computer Engineering,

Instituto Superior Técnico, Lisboa, Portugal

1. Genetic information is stored in DNA, which is essentially a long sequence of symbols from a 4-letter alphabet $\{A, C, T, G\}$ (standing for the four DNA bases: Adenine, Cytosine, Thymine, and Guanine).

   a) If we interpret a DNA sequence as generated randomly by a memoryless source (which, of course, it is not) with uniform distribution, what is the entropy (in base-2, that is, expressed in bits/symbol) of this source?

   b) Repeat the previous question, but using base-4 entropy.

   c) Knowing that the length of human DNA sequence is 3.2×10^9 bases, how many bits are required to store its information? Compare the result with common storage media (for example, USB flash drive of common size).

2. Consider four random variables, $X \in \{1, 2\}$, $Y \in \{1, 2, 3\}$, $Z \in \{1, 2, 3\}$, $T \in \{1, 2, 3, 4, 5\}$, with the following probability distributions

$P(\cdot)$	1	2	3	4	5
X	1/2	1/2	·	·	·
Y	1/3	1/3	1/3	·	·
Z	1/2	1/4	1/4	·	·
T	1/3	1/4	1/4	1/12	1/12

   a) Without computing the entropies, sort the four variables by increasing order of uncertainty.

   b) Find the entropies of the variables and confirm the order obtained in the previous question.

3. Let the random variable X correspond to the sum of the outcomes of a pair of fair dice, denoted Y and Z, that is $X = Y + Z$.

   a) Compute the entropies $H(Y)$, $H(Z)$, and $H(Y, Z)$.

   b) Find the probability distribution of X and the corresponding entropy $H(X)$ (with base 2).

4. Consider three binary variables such that (X_1, X_2, X_3) take the following values with equal probability: $(0, 0, 0)$, $(0, 1, 0)$, $(1, 0, 0)$, and $(1, 0, 1)$. Use the chain rule for entropy to compute $H(X_1, X_2, X_3)$ (which of course is 2 bits/symbol).

5. Consider two variables $X, Y \in \{1, 2, 3\}$ with the following joint pmf:

		Y		
		1	2	3
X	1	1/6	1/6	0
	2	0	1/6	1/6
	3	1/6	0	1/6

a) Compute the entropies $H(X)$ and $H(Y)$, the joint entropy $H(X,Y)$, and the conditional entropies $H(Y)|X)$ and $H(X|Y)$.

b) Repeat the previous question, with the following joint pmfs:

		Y		
		1	2	3
X	1	1/2	0	0
	2	0	1/4	0
	3	0	1/4	0

and

		Y		
		1	2	3
X	1	1/6	1/12	1/12
	2	1/3	1/6	1/6

6. Consider a pair of binary random variables $(X,Y) \in \{(0,0),(0,1),(1,0),(1,1)\}$, with the following joint probability distribution: $P(X=0,Y=0)=1/3$, $P(X=0,Y=1)=1/3$, $P(X=1,Y=0)=0$, $P(X=1,Y=1)=1/3$. Find $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $H(X,Y)$ and $I(X;Y)$.

7. Let X be a random variable with values in a set $\mathcal{X} = \{x_1,...,x_N\}$. Consider a deterministic function $f : \mathcal{X} \to \mathcal{Y}$.

   a) Show that $H(f(X)) \leq H(X)$.

   b) Give examples where $H(f(X)) = H(X)$ and $H(f(X)) < H(X)$.

   c) Obtain an expression for $I(X, f(X))$

8. Let X be a random variable taking values in the following set of integers: $\mathcal{X} = \{-3,-2,-1, 0,1,2,3\}$.

   a) Let $Y = X^3$; what is the relationship between $H(Y)$ and $H(X)$ (valid for any probability distribution)?

   b) Let $Y = X^2$; what is the relationship between $H(Y)$ and $H(X)$ (valid for any probability distribution)? Can we have $H(Y) = H(X)$? Justify and give examples.

9. Consider two random variables X and Y, taking values, respectively in $\mathcal{X} = \{x_1,...,x_N\}$ and $\mathcal{Y} = \{y_1,...,y_M\}$. Assume that $\mathcal{X} \subset \mathbb{R}$ and $\mathcal{Y} \subset \mathbb{R}$, that is, the elements of $\mathcal{X}$ and $\mathcal{Y}$ are real numbers. Finally, consider the random variable $Z = X + Y$.

   a) Show that $H(Z|X) = H(Y|X)$.

   b) Show that if X and Y are independent, then $H(Z) \geq H(X)$ and $H(Z) \geq H(Y)$; that is, show that adding independent variables increases the entropy.

   c) Give examples of cases (necessarily involving non-independent variables X and Y) where $H(Z) < H(X)$ and $H(Z) < H(Y)$.

   d) Give an exemple where $H(Z) = H(X) + H(Y)$.