Information and Coding Theory

Exercise Sheet 2

Exercise 1. The random variables X and Y have the joint probability distribution given by the table below:

X \ Y	0	1	2
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
1	$\frac{1}{9}$	0	$\frac{2}{9}$
2	$\frac{1}{9}$	$\frac{2}{9}$	0

Find:

(a) $H(X)$ and $H(Y)$.

(b) $H(X,Y)$.

(c) $H(X|Y)$ and $H(Y|X)$.

(d) $I(X;Y)$.

(e) Draw a Venn diagram for the quantities obtained above.

Exercise 2. An urn contains r red, g green, and b black balls. We consider as a random variable X the color of a ball that is randomly drawn from the urn.

(a) Show that the entropy of drawing k balls with replacement (i.e., putting the ball back to the urn after retrieving its color) is given by $H(X_1,X_2,...,X_k) = k \cdot H(X)$ where $H(X)$ is the entropy of a single draw.

(b) The balls are not replaced, and we do not look at the color of the balls that were taken out during the first $i-1$ draws (or, in other words, the colors of the preceding $i-1$ balls are forgotten). Give an intuitive argument why $H(X_i) = H(X)$ where $H(X)$ is the entropy of an arbitrary draw and X_i is the color of the ball at the ith draw. Without appealing to intuition, the rigorous proof of this claim follows after solving (c) to (f).

(c) The balls are not replaced. Compare the probability of having the colors c_1 at the first draw and c_2 at the second draw to the probability of having the color c_2 at the first draw and c_1 at the second draw.

(d) The balls are not replaced. What is the probability that a red ball is drawn at the second draw?

(e) The balls are not replaced. What can we say about the marginal probabilities of drawing a specific color at the second draw?

(f) The balls are not replaced. Compare the entropy of the second draw with that of the first draw.

(g) Show that the entropy of drawing k balls without replacements is less than or equal to the entropy of drawing k balls with replacement.

Exercise 3. The random variables X, Y and Z form a Markov chain $X \to Y \to Z$ if the conditional probability of Z depends only on Y, i.e. $p(X,Y,Z) = p(X)p(Y|X)p(Z|Y)$.

(a) Show that X and Z are conditionally independent given Y, i.e. $p(X,Z|Y) = p(X|Y)p(Z|Y)$.

(b) Show that $I(X;Y) \geq I(X;Z)$ (data processing inequality) by using the result of (a) and the chain rule for mutual information $I(X;Y|Z)$.

(c) The random variables X, Y and Z have probability distributions on sets $\mathcal{X}$ of cardinality n, $\mathcal{Y}$ of cardinality k and $\mathcal{Z}$ of cardinality m. Show that if $n > k$ and $m > k$ then $I(X;Z) \leq \log_2 k$.

(d) Explain what happens if $k = 1$.

Exercise 4. Consider binary sequences of length n emitted by a source of independent and identically distributed (i.i.d.) random variables. Each bit is emitted independently of the previous one according to the binomial distribution where with probability p a 1 is emitted. A **typical sequence** contains k ones and $(n-k)$ zeros, with $k \simeq np$ (the mean number of ones is rounded to the closest integer.)

(a) What is the probability that the source emits a particular typical sequence? Give an approximation to that probability as a function of the entropy of the source.

(b) What is the number of typical sequences, expressed as a function of the entropy of the source? Compare it with the absolute number of sequences that can be emitted by the source. What is the approximate probability that a typical sequence is emitted? (Use the Stirling approximation $n! \approx n^n$)

(c) What is the most probable sequence that can be emitted by the source? Is it typical?

Exercise 5.

Definition. A real-valued map $\rho(x,y)$ is a **metric** (or distance function) if it satisfies the following properties $\forall x,y,z$:

- $\rho(x,y) \geq 0$
- $\rho(x,y) = \rho(y,x)$
- $\rho(x,y) = 0$ if and only if $x = y$
- $\rho(x,y) + \rho(y,z) \geq \rho(x,z)$

(a) Let $\rho(X,Y) = H(X|Y) + H(Y|X)$. Verify that $\rho(X,Y) = H(X,Y) - I(X;Y) = 2H(X,Y) - H(X) - H(Y)$, and interpret this relation with the help of a Venn diagram.

(b) Prove that $\rho(X,Y)$ satisfies all properties of a metric if the notation $X = Y$ means that there exists a bijection between X and Y.

Hint: To show the last property, use the strong subadditivity.

Exercise 6. Let X, Y and Z be uniformly distributed binary random variables such that $H(X) = H(Y) = H(Z) = 1$ bit. If Z is ignored, the mutual information of X and Y can be defined as $I(X;Y) = H(X) - H(X|Y)$. If Z is taken into account, the mutual information, **conditioned** on Z, can be defined as $I(X;Y|Z) = H(X,Z) - H(X|Y,Z)$. Finally, the mutual information of all three random variables can be defined as $I(X;Y;Z) = I(X;Y) - I(X;Y|Z)$. The last quantity can be positive, negative, or zero.

(a) Give an example of a distribution of X, Y, and Z, such that $I(X;Y;Z) > 0$.

(b) Give an example of a distribution of X, Y, and Z, such that $I(X;Y;Z) < 0$.

The exercises and solutions are available at `http://quic.ulb.ac.be/teaching`