Information Theory Exercise Sheet #6

University of Amsterdam, Master of Logic, Fall 2014
Lecturer: Christian Schaffner
TA: Philip Schulz

Out: Wednesday, 3 December 2014 2014
(due: Friday, 12 December 2014, 11:00)

Previous Exam Questions

1. In this exercise we consider yet another different entropy notion. Let X and Y be random variables with joint probability distribution P_{XY}. The *collision probability* and the *collision entropy* are respectively defined as

$$\mathrm{Col}(X) := \sum_x P_X(x)^2 \quad \text{and} \quad H_2(X) := -\log \mathrm{Col}(X).$$

The *conditional collision probability* and the *conditional collision entropy* are respectively defined as

$$\mathrm{Col}(X|Y) := \sum_y P_Y(y)\mathrm{Col}(X|Y=y) \quad \text{and} \quad H_2(X|Y) := -\log \mathrm{Col}(X|Y).$$

(a) Prove that $H_2(X) \leq H_2(XY)$.

(b) Prove that $H_2(X|Y) \leq H_2(X)$.

(c) Prove that

$$0 \leq H_{\min}(X) \leq H_2(X) \leq H(X)$$

and

$$0 \leq H_{\min}(X|Y) \leq H_2(X|Y) \leq H(X|Y)\,.$$

To be solved in Class

1. Prove Lemma 1 below stating that the capacity per transmission is not increased if we use a discrete memoryless channel many times. For inspiration, look again at the proof of the converse of Shannon's noisy-channel coding theorem.

Lemma 1 (Lemma 7.9.2 in [CT]) *Let Y^n be the result of passing X^n through a discrete memoryless channel of capacity C. Then, $I(X^n; Y^n) \leq nC$ for all P_{X^n}.*

Does your proof also work for the feedback case (i.e. where X_{i+1} is allowed to depend on X^iY^i)? If not, point out the steps in your proof where you use that there is no feedback.

2. *Symmetric Channels.* Consider the channel with transition matrix

$$P_{Y|X} = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}.$$

In a transition matrix, the entry in the xth row and yth column denotes the conditional probability $P_{Y|X}(y|x)$ that y is received when x has been sent.

Definition 1 *A channel is said to be* symmetric *if the rows of the channel transition matrix $P_{Y|X}$ are permutations of each other and the columns are permutations of each other. A channel is said to be* weakly symmetric *if every row of the transition matrix is a permutation of every other row and all the column sums $\sum_x P_{Y|X}(y|x)$ are equal.*

For instance, the channel $P_{Y|X}$ above is symmetric and the channel

$$Q_{Y|X} = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{bmatrix}$$

is weakly symmetric but not symmetric.

(a) Find the optimal input distribution and channel capacity of $Q_{Y|X}$.

(b) Give a general strategy how to compute the capacity for weakly symmetric channels. What is the optimal input distribution?

3. *Geometric distribution.* For a $0 \leq p \leq 1$, let us consider a series of independent events that each have success probability p. Let X be the number of trials until the first success.

(a) Show that $P_X(n) = (1-p)^{n-1}p$.

(b) Give closed formulas for $\sum_{n=0}^{\infty} np^n$ and $\sum_{n=0}^{\infty} n^2 p^n$.
Hint: Recall the formula for a geometric series $\sum_{n=0}^{\infty} p^n = \frac{1}{1-p}$. Differentiate with respect to p on both sides.

(c) Show that the entropy $H(X)$ is $\frac{h(p)}{p}$.

(d) Compute $\mathbb{E}[X]$.

(e) Compute $\mathrm{Var}[X]$.

Homework

1. *Zero-error vs non-zero-error Shannon capacity:* Let $P_{Y|X}$ be a discrete memoryless channel with confusability graph G and capacity $C = \max_{P_X} I(X;Y)$.

(a) [2 points] Show that $\log(\alpha(G)) \leq C$.

(b) [2 points] Show that for any $n \geq 1$, $\log(\alpha(G^{\boxtimes n})) \leq \max_{P_{X^n}} I(X^n;Y^n)$, where the Y^n are obtained by using the channel n times, i.e. $P_{Y^n|X^n}(y^n|x^n) = \Pi_{i=1}^n P_{Y|X}(y_i|x_i)$ for all x^n, y^n.

(c) [2 points] Conclude that the zero-error Shannon capacity of G is at most the channel capacity C.

2. [6 points] *Additive noise channel.* Find the channel capacity of the following discrete memoryless channel. On input X from $\mathcal{X} = \{0,1\}$, the output Y is obtained by adding (over the reals) another real random variable Z, i.e. $Y = X + Z$ with distribution $P_Z(0) = P_Z(a) = \frac{1}{2}$ independent of X. Compute the channel capacity for all possible values of $a \in \mathbb{R}$.

3. *Tall, fat people.* Suppose that the average height of people in a room is 1.5m. Suppose that the average weight is 50kg.

   (a) [1 point] Argue that no more than one third of the population is 4.5m tall.

   (b) [2 points] Find an upper bound on the fraction of people who are simultanously tall (say, at least 3m) and fat (say, at least 150kg).

4. *Another Kind of Entropy.* In this exercise we consider a different entropy notion. Let X and Y be random variables with joint probability distribution P_{XY}. The *guessing probability* and the *min-entropy* of X are respectively defined as

$$\text{Guess}(X) := \max_x P_X(x) \quad \text{and} \quad H_{\min}(X) := -\log \text{Guess}(X).$$

   The *conditional guessing probability* and the *conditional min-entropy* of X are respectively defined as

$$\text{Guess}(X|Y) := \sum_y P_Y(y)\text{Guess}(X|Y=y)$$

   and

$$H_{\min}(X|Y) := -\log \text{Guess}(X|Y).$$

   (a) [1 point] If X has no uncertainty (i.e. $H(X)=0$), what is $H_{\min}(X)$?

   (b) [1 point] If X is uniformly distributed over $\mathcal{X}$, what is $H_{\min}(X)$?

   (c) [2 points] Prove that $H_{\min}(XY) \geq H_{\min}(X)$.

   (d) [2 points] Prove that $H_{\min}(X) \geq H_{\min}(X|Y)$.

   (e) [2 points] Prove that $H_{\min}(X|Y) \geq H_{\min}(XY) - \log|\mathcal{Y}|$.

5. *Erasures and errors in a binary channel* Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be ε and the probability of erasure be α, so the channel is as described in Figure 1.

   (a) [3 points] Find the channel capacity of this channel.

   (b) [1 point] Specialize to the case of the binary symmetric channel ($\alpha = 0$).

   (c) [1 point] Specialize to the case of the binary erasure channel ($\varepsilon = 0$).

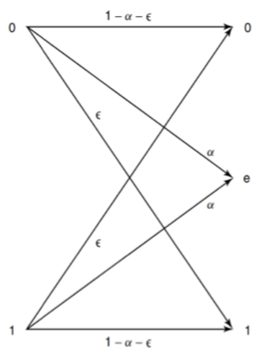

Figure 1: Erasures and errors in a binary channel.

6. *Encoder and decoder as part of the channel.* Consider a binary symmetric channel (BSC) $P_{Y|X}$ with crossover probability $\varepsilon = 0.1$. A possible coding scheme for this channel with two codewords of length 3 is to encode message w_1 as 000 and w_2 as 111. The decoder uses majority vote. With this coding scheme, we can consider the combination of encoder, channel, and decoder as forming a new BSC $Q_{Y|X}$, with two inputs w_1 and w_2 and two outputs w_1 and w_2.

   (a) [3 points] Calculate the crossover probability of this new channel $Q_{Y|X}$.

   (b) [2 points] What is the capacity of this new channel in bits per transmission of the original channel $P_{Y|X}$?

   (c) [1 point] What is the capacity of the original BSC $P_{Y|X}$ with crossover probability $\varepsilon = 0.1$. Compare the two capacities.

   (d) [4 points] Prove the following general result: For any channel, considering the encoder, channel, and decoder together (as a new channel from message W to estimated messages $\hat{W}$) will not increase the capacity in bits per transmission of the original channel.