Information Theory Exercise Sheet #3

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

Out: Wednesday, 12 November 2014 2014
(due: Wednesday, 19 November 2014, 13:00)

Previous Exam Questions

At this point in the course, you should be able to solve the following exercises which were exam questions in previous editions of the course:

1. Let X, Y, Z be *binary* random variables such that $I(X;Y) = 0$ and $I(X;Z) = 0$.
   (a) [$\frac{1}{2}$ points] Does it follow that $I(X;Y,Z) = 0$? If yes, prove it. If no, give a counterexample.
   (b) [$\frac{1}{2}$ points] Does it follow that $I(Y;Z) = 0$? If yes, prove it. If no, give a counterexample.

2. Let A, B, C be random variables over alphabet $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ for some integer $n \geq 2$. Let us assume that

$$A = B + C \mod n\,, \tag{1}$$
$$H(B) = \log(n)\,, \tag{2}$$
$$I(A;B) = 0\,. \tag{3}$$

   Show that $I(A;C) = 0$.

3. Let A, B, C be random variables such that

$$I(A;B) = 0\,, \tag{4}$$
$$I(A;C|B) = I(A;B|C)\,, \tag{5}$$
$$H(A|BC) = 0\,. \tag{6}$$

   What is the relation between the quantities $H(A)$ and $H(C)$?

To be solved in Class

1. For the Markov chain $X \leftrightarrow Y \leftrightarrow \hat{X}$, show that $H(X|\hat{X}) \geq H(X|Y)$.
2. [Cover-Thomas 2.32]. We are given the following joint distribution of $X \in \{1, 2, 3\}$ and $Y \in \{a, b, c\}$:

$$P_{XY}(1, a) = P_{XY}(2, b) = P_{XY}(3, c) = 1/6$$
$$P_{XY}(1, b) = P_{XY}(1, c) = P_{XY}(2, a) = P_{XY}(2, c) = P_{XY}(3, a) = P_{XY}(3, b) = 1/12.$$

   Let $\hat{X}(Y)$ be an estimator for X (based on Y) and let $p_e = P(\hat{X} \neq X)$.

   (a) Find an estimator $\hat{X}(Y)$ for which the probability of error p_e is as small as possible.

   (b) Evaluate Fano's inequality for this problem and compare.
3. The mean of a random variable X is $\mu = \mathbb{E}[X]$. The variance of X is defined as $\mathrm{Var}[X] = \mathbb{E}[(X - \mu)^2]$.

   (a) Show that $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

   (b) Show that for any real $a > 0$, it holds that $\mathrm{Var}[aX] = a^2 \mathrm{Var}[X]$, and $\mathrm{Var}[X + a] = \mathrm{Var}[X]$.

   (c) Show that for independent random variables X, Y, we have $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

   (d) Let X be a random variable with *Bernoulli distribution* $P_X(1) = p$ and $P_X(0) = 1 - p$. Compute $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

   (e) Let Y be a random variable with *binomial distribution* $P_Y(y) = \binom{n}{y} p^y (1 - p)^{n-y}$. Compute $\mathbb{E}[Y]$ and $\mathrm{Var}[Y]$.

Homework

1. *Deriving the weak law of large numbers.*

   (a) [3 points] (Markov's inequality.) For any real non-negative random variable X and any $t > 0$, show that

$$P_X(X \geq t) \leq \frac{\mathbb{E}[X]}{t} .$$

   Exhibit a random variable (which can depend on t) that achieves this inequality with equality.

   (b) [2 points] (Chebyshev's inequality.) Let Y be a random variable with mean μ and variance σ^2. By letting $X = (Y - \mu)^2$, show that for any $\varepsilon > 0$,

$$P\left(|Y - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2} .$$

   (c) [2 points] (The weak law of large numbers.) Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of iid random variables with mean μ and variance σ^2. Let $\overline{Z_n} = \frac{1}{n} \sum_{i=1}^{n} Z_i$ be the sample mean. Show that

$$P\left(\left|\overline{Z_n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} .$$

   Thus, $P\left(\left|\overline{Z_n} - \mu\right) > \varepsilon\right) \to 0$ as $n \to \infty$. This is known as the weak law of large numbers.

2. *AEP and source coding.* A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $P_X(1) = 0.005$ and $P_X(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer 1's.

   (a) [2 points] Assuming that all codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer 1's.

   (b) [2 points] Calculate the probability of observing a source sequence for which no codeword has been assigned.

   (c) [3 points] Use Chebychev's inequality to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part (b).

3. *Calculation of typical set.* To clarify the notion of a typical set $A_\varepsilon^{(n)}$ and the smallest set of high probability $B_\delta^{(n)}$, we will calculate these sets for a simple example. Consider a sequence of iid binary random variables $X_1, X_2, \dots X_n$, where the probability that $P_X(1) = 0.6$ and $P_X(0) = 0.4$.

   (a) [1 point] Calculate H(X).

   (b) [3 points] With $n = 25$ and $\varepsilon = 0.1$, which sequences fall in the typical set $A_\varepsilon^{(n)}$? What is the probability of the typical set? How many elements are there in the typical set? (This involves computation of a table of probabilities for sequences with k 1's, $0 \leq k \leq 25$, and finding those sequences that are in the typical set.)

   Hint: Here is the table: `http://goo.gl/sQCPMO`

   (c) [2 points] How many elements are there in the smallest set that has probability 0.9? In other words, what is $|B_\delta^{(n)}|$ for $n = 25$ and $\delta = 0.1$?

   (d) [2 points] How many elements are there in the intersection $|A_\varepsilon^{(n)} \cap B_\delta^{(n)}|$ of the sets computed in parts (b) and (c)? What is the probability of this intersection?