Information Theory Exercise Sheet #3

University of Amsterdam, Master of Logic, Spring 2014
Lecturer: Christian Schaffner
TA: Hoang Cuong
Out: Thu, 20 February 2014
Due: Thu, 27 February 2014, 11:00

To be solved in Class

1. The mean of a random variable $X$ is $\mu = \mathbb{E}[X]$. The variance of $X$ is defined as $\text{Var}[X] = \mathbb{E}[(X - \mu)^2]$.

(a) Show that $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
(b) Show that for any real $a > 0$, it holds that $\text{Var}[aX] = a^2 \text{Var}[X]$, and $\text{Var}[X + a] = \text{Var}[X]$.
(c) Show that for independent random variables $X, Y$, we have $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$.
(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 $\text{Var}[X]$.
(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 $\text{Var}[Y]$.

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 $\mu$ and variance $\sigma^2$. By letting $X = (Y - \mu)^2$, show that for any $\varepsilon > 0$,
   $$P(|Y - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$  
(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 $\mu$ and variance $\sigma^2$. Let $\overline{Z_n} = \frac{1}{n} \sum_{i=1}^{n} Z_i$ be the sample mean. Show that
   $$P(|\overline{Z_n} - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$  
   Thus, $P(|\overline{Z_n} - \mu| > \varepsilon) \to 0$ 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 \( X_1, X_2, \ldots, X_n \), 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/sQCPM0](http://goo.gl/sQCPM0)

(c) [2 points] How many elements are there in the smallest set that has probability 0.9?

(d) [2 points] How many elements are there in the intersection of the sets in parts (c) and (d)? What is the probability of this intersection?