Information Theory I

Fall Semester 2016

Prof. Dr. A. Lapidoth

Signal and Information Processing Laboratory

Institut für Signal- und Informationsverarbeitung

Exercise 4 of October 12, 2016

http://www.isi.ee.ethz.ch/teaching/courses/it1/

Problem 1 — *Fano's Inequality*

Let X take values in the set $\mathcal{X} = \{1, \ldots, m\}$ and let $p_i = \Pr[X = i]$ for $i = 1, \ldots, m$.

a) Suppose that you must guess X. Which guess would have the smallest probability of error? What is the probability of error P_e^* associated with this guess?

b) Maximize $H(X)$ over all choices of $p_1, p_2, \ldots, p_m$ for which the best guess has a probability of error P_e^*, where P_e^* is a constant in the interval $\left[0, \frac{m-1}{m}\right]$.

c) Suppose that you must guess X based on the observation of a correlated chance variable Y. Given some guessing rule, let P_e be the probability of error and let E be a chance variable which is zero if the guess is correct and one otherwise. Justify the following proof of Fano's Inequality:

$$\begin{aligned} H(X|Y) &\overset{(i)}{=} \sum_{y\in\mathcal{Y}} P_Y(y) H(X|Y=y) \\ &\overset{(ii)}{\leq} \sum_{y\in\mathcal{Y}} P_Y(y) \Big[H(E|Y=y) + \Pr[E=1|Y=y]\log(m-1) \Big] \\ &\overset{(iii)}{=} H(E|Y) + \Pr[E=1]\log(m-1) \\ &\overset{(iv)}{\leq} H(E) + \Pr[E=1]\log(m-1) \\ &\overset{(v)}{=} H_b(P_\text{e}) + P_\text{e}\log(m-1). \end{aligned}$$

Hint: Use the result from Part b).

Problem 2 — *Slackness in Kraft's Inequality*

Let $l_1, l_2, \ldots, l_m$ denote the codeword lengths of an instantaneous code of the D-ary alphabet $\mathcal{D} = \{0, 1, \ldots, \mathrm{D}-1\}$.

a) For $\mathrm{D} = 2$, i.e. binary codes, show that if

$$\sum_{i=1}^{m} 2^{-l_i} < 1,$$

then there exists another instantaneous binary code that is deterministically better than the original code. Deterministically better means that for each source symbol, the corresponding codeword is no longer than the codeword from the original code, and that for at least one source symbol, the corresponding codeword is shorter.

b) For $\mathsf{D} > 2$, (e.g. ternary codes, quaternary codes, ...) with

$$\sum_{i=1}^{m} \mathsf{D}^{-l_i} < 1,$$

does the same conclusion hold?

Problem 3 — *Optimal Code Lengths that Require One Bit above Entropy*

The source coding theorem asserts that an optimal code for a chance variable X has expected length less than $H(X)+1$. Show that this upper bound cannot be improved by constructing, for every $\epsilon > 0$, a chance variable for which the least expected length is larger than $H(X)+1-\epsilon$.

Problem 4 — *Shannon Code*

Consider the following method for generating a code for a random variable X which takes on m values $\{1, 2, \ldots, m\}$ with positive probabilities $p_1, p_2, \ldots, p_m$. Assume that the probabilities are ordered so that $p_1 \geq p_2 \geq \cdots \geq p_m$. Define

$$\mathsf{F}_i = \sum_{k=1}^{i-1} p_k,$$

the sum of the probabilities of all symbols less than i. Then the codeword for i is the binary representation[1] of the number $\mathsf{F}_i \in [0, 1)$ truncated to l_i bits, where $l_i = \lceil \log \frac{1}{p_i} \rceil$.

a) Show that the code constructed by this process is prefix-free and the average length satisfies

$$H(X) \leq \mathsf{L} < H(X) + 1.$$

b) Construct the code for the probability distribution $(0.5, 0.25, 0.125, 0.125)$.

[1] In cases where the binary representation is not unique we always take the terminating expansion. For example, we represent $1/2$ by 0.1 instead of $0.0111\ldots$