# Information Theory I

Fall Semester 2016

Prof. Dr. A. Lapidoth

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

## Model Answers to Exercise 4 of October 12, 2016

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

---

### Problem 1 *Fano's Inequality*

a) For any guess $\hat{x} \in \mathcal{X}$, the probability of error is

$$P_{\mathrm{e}} = \Pr[X \neq \hat{x}] = \sum_{x \in \mathcal{X} \setminus \{\hat{x}\}} P_X(x) = 1 - P_X(\hat{x}),$$

so the smallest probability of error is achieved by guessing the most probable value of $X$ (or any of the most probable values if the most probable value is not unique). The associated probability of error is

$$P_{\mathrm{e}}^* = 1 - \max_{x \in \mathcal{X}} P_X(x).$$

b) From Part a) we know that the most probable value of $X$ must have probability $1 - P_{\mathrm{e}}^*$, so without loss of generality let $p_1 = 1 - P_{\mathrm{e}}^*$. To maximize $H(X)$ subject to $p_1 = 1 - P_{\mathrm{e}}^*$, we use the fact that the entropy is *concave*. Let $(1 - P_{\mathrm{e}}^*, p_2, \ldots, p_m)$ be a probability distribution that maximizes $H(X)$ and denote that maximum by $H^*$. Because relabeling does not change the entropy, the entropies of the following probability distributions are also equal to $H^*$:

$$\begin{aligned}
P_2 &= (1 - P_{\mathrm{e}}^*,\ p_2,\ p_3,\ \ldots,\ p_{m-1}, p_m),\\
P_3 &= (1 - P_{\mathrm{e}}^*,\ p_3,\ p_4,\ \ldots,\ p_m,\ \ p_2),\\
&\vdots\\
P_m &= (1 - P_{\mathrm{e}}^*, p_m, p_2,\ \ldots,\ p_{m-2}, p_{m-1}).
\end{aligned}$$

Consequently,

$$\begin{aligned}
H^* &\overset{\text{(i)}}{\geq} H\left(\frac{1}{m-1}P_2 + \frac{1}{m-1}P_3 + \ldots + \frac{1}{m-1}P_m\right)\\
&\overset{\text{(ii)}}{\geq} \frac{1}{m-1}H(P_2) + \frac{1}{m-1}H(P_3) + \ldots + \frac{1}{m-1}H(P_m)\\
&= H^*,
\end{aligned}$$

where (i) follows because $H^*$ is the maximum entropy subject to $p_1 = 1 - P_{\mathrm{e}}^*$ and (ii) follows from the concavity of the entropy. Because $p_2 + p_3 + \ldots + p_m = 1 - p_1 = P_{\mathrm{e}}^*$, the mixture $\frac{1}{m-1}P_2 + \frac{1}{m-1}P_3 + \ldots + \frac{1}{m-1}P_m$ is equal to $(1 - P_{\mathrm{e}}^*, \frac{P_{\mathrm{e}}^*}{m-1}, \frac{P_{\mathrm{e}}^*}{m-1}, \ldots, \frac{P_{\mathrm{e}}^*}{m-1}, \frac{P_{\mathrm{e}}^*}{m-1})$, and

$$\begin{aligned}
H^* &= H\left(\frac{1}{m-1}P_2 + \frac{1}{m-1}P_3 + \ldots + \frac{1}{m-1}P_m\right)\\
&= (1 - P_{\mathrm{e}}^*)\log\frac{1}{1 - P_{\mathrm{e}}^*} + (m-1)\frac{P_{\mathrm{e}}^*}{m-1}\log\frac{m-1}{P_{\mathrm{e}}^*}\\
&= H_b(P_{\mathrm{e}}^*) + P_{\mathrm{e}}^*\log(m-1).
\end{aligned}$$

c) i) This follows from the definition of the conditional entropy.

ii) Since the guess is based on the observation of $Y$, we are interested in the distribution of $X$ conditional on $Y = y$. For this conditional distribution, we infer from Part b) that $H(X|Y = y) \leq H_b(P_{e,y}) + P_{e,y} \log(m-1)$, where $P_{e,y}$ denotes the probability of error conditional on $Y = y$. Rewriting this inequality in terms of the error indicator variable $E$, we obtain $H(X|Y = y) \leq H(E|Y = y) + \Pr[E = 1|Y = y] \log(m-1)$ for every $y \in \mathcal{Y}$.

iii) This follows from the definition of the conditional entropy and from the law of total probability.

iv) This follows because conditioning does not increase entropy.

v) This follows because $E$ is a binary random variable and $\Pr[E = 1] = P_e$.

## Problem 2 *Slackness in Kraft's Inequality*

a) The codewords of an instantaneous code can be assigned to internal nodes or leaves of a binary tree of depth $l_{\max} = \max\{l_1, l_2, \ldots, l_m\}$ such that a codeword of length $l_i$ is assigned to an internal node at depth $l_i$ and all the children of this node cannot represent any other codeword. We will describe this later fact by saying that the $2^{l_{\max}-l_i}$ descending leaves are shadowed, which is depicted in Fig. 1 i) using dotted lines and gray colored leaves. (In case the node representing the codeword is a leaf itself, we will also say that the leaf is shadowed.)

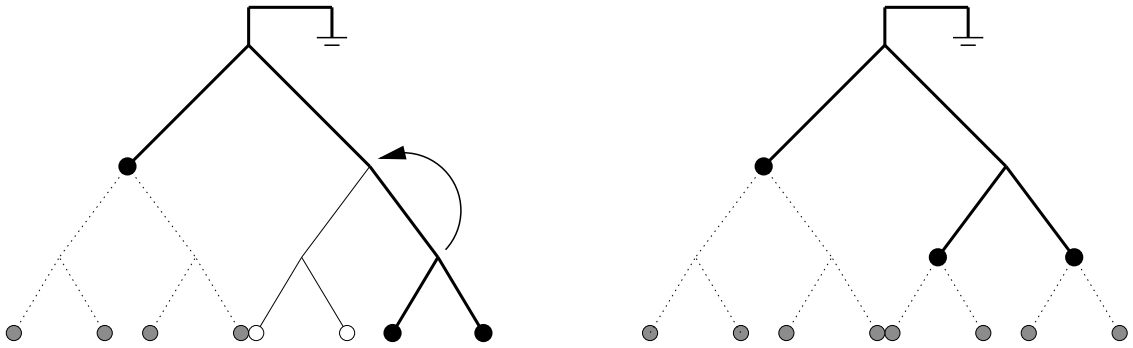


Figure 1: A tree representing a prefix-free code satisfying Kraft's inequality with strict inequality is depicted in i), and a tree corresponding to a better prefix-free code satisfying Kraft's inequality with equality is shown in ii). Nodes corresponding to codewords are black, shadowed leaves (unless they correspond to codewords) are gray and unshadowed leaves are white.

The tree consists of $2^{l_{\max}}$ leaves, and the total number of leaves which are shadowed is

$$\sum_{i=1}^{m} 2^{l_{\max}-l_i} = 2^{l_{\max}} \sum_{i=1}^{m} 2^{-l_i} \overset{(*)}{<} 2^{l_{\max}} \cdot 1 = 2^{l_{\max}},$$

where the strict inequality $(*)$ holds by assumption. Hence there are leaves which are not shadowed by any codeword, i.e., on the path from the leave to the root there is no codeword assigned. Consequently we can improve our code in the following way:

 

- choose one of the leaves that is not shadowed and follow its path up to the root as long as the current node has only descendant leaves which are not shadowed;
- as soon as you encounter an internal node for which a part of the descendants is shadowed, remove that internal node and replace it by the child which belongs to the path leading to the shadowed leaf.

(For an illustration see the transition from Fig. 1 i) to ii).) This clearly reduces the length of the codewords.

b) The conclusion generally does not hold for $\mathsf{D} > 2$. A simple example is to choose $\mathsf{D} = 3$ and $m = 2$. Clearly, the "best" ternary code for a source with 2 possible outcomes uses one symbol to describe each outcome. Then we have $l_1 = l_2 = 1$ and

$$\sum_{i=1}^{2} \mathsf{D}^{-l_i} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} < 1.$$

Thus for our code Kraft's inequality holds with strict inequality. Nevertheless, it is not possible to find a deterministically better code for the described source.

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

Let $X$ be a binary chance variable with probability mass function

$$P_X(x) = \begin{cases} 1-\delta & \text{for } x = x_0, \\ \delta & \text{for } x = x_1, \end{cases}$$

with $0 < \delta < 1$. Then an optimal code assigns to both symbols a codeword of length 1, e.g.,

$$c(x_0) = 0,$$
$$c(x_1) = 1,$$

where $c(x)$ denotes the codeword for $x$. The expected length of this code is

$$\mathsf{L} = 1.$$

However, the entropy of $X$ is $H(X) = H_{\mathrm{b}}(\delta)$ where $H_{\mathrm{b}}(\cdot)$ denotes the binary entropy function. Since $\inf_\delta H_{\mathrm{b}}(\delta) = 0$ it follows that for any $\epsilon > 0$ there exists a $\delta \in (0,1)$ such that $H_{\mathrm{b}}(\delta) < \epsilon$. Thus

$$\mathsf{L} = 1 > 1 + H_{\mathrm{b}}(\delta) - \epsilon = H(X) + 1 - \epsilon.$$

## Problem 4 *Shannon Code*

a) Let $c_i$ be the codeword corresponding to $\mathsf{F}_i$. In order to show that the code is prefix-free, we first note that since $p_1 \geq p_2 \geq \cdots \geq p_m$, we have $l_1 \leq l_2 \leq \cdots \leq l_m$. We further note that for any $j > i$,

$$\mathsf{F}_j - \mathsf{F}_i = \sum_{k=i}^{j-1} p_k \geq \sum_{k=i}^{j-1} 2^{-l_k}, \tag{1}$$

where the inequality follows because $l_k = \left\lceil \log \frac{1}{p_k} \right\rceil \geq \log \frac{1}{p_k}$.

In the following we show that for $j > i$ and thus $l_i \leq l_j$ the codeword $c_i$ cannot be a prefix of $c_j$. To this end, note that $c_i$ can only be a prefix of $c_j$, if the first $l_i$ symbols of $c_j$ are identical to $c_i$, or equivalently if the first $l_i$ bits in the binary representation of $\mathsf{F}_j$ are the same as the first $l_i$ bits in the binary representation of $\mathsf{F}_i$. But this can only be the case if $\mathsf{F}_j - \mathsf{F}_i < 2^{-l_i}$ which by (1) is not possible, and consequently, $c_i$ can never be a prefix of $c_j$. This concludes the proof.

Now we look at the expected length:

$$\begin{aligned}
\log\frac{1}{p_i} &\leq \quad l_i \quad &&< \log\frac{1}{p_i} + 1,\\
p_i \log\frac{1}{p_i} &\leq \quad p_i l_i \quad &&< p_i \log\frac{1}{p_i} + p_i,\\
\sum_{i=1}^{m} p_i \log\frac{1}{p_i} &\leq \sum_{i=1}^{m} p_i l_i &&< \sum_{i=1}^{m} p_i \log\frac{1}{p_i} + \sum_{i=1}^{m} p_i,\\
H(X) &\leq \quad \mathsf{L} \quad &&< H(X) + 1.
\end{aligned}$$

b)

$$\begin{aligned}
&\mathsf{F}_1 = 0 && \equiv 0.0000 && l_1 = 1 \implies 0,\\
&\mathsf{F}_2 = 0.5 && \equiv 0.1000 && l_2 = 2 \implies 10,\\
&\mathsf{F}_3 = 0.75 && \equiv 0.1100 && l_3 = 3 \implies 110,\\
&\mathsf{F}_4 = 0.875 && \equiv 0.1110 && l_4 = 3 \implies 111.
\end{aligned}$$

The codewords are then $\{0, 10, 110, 111\}$.