Information Theory and Coding | EPFL Winter Semester 2008/2009
Prof. Suhas Diggavi | Handout #5, Wednesday, 1 October 2008

Homework Set #2

Due 9 October 2008 (Before 14:00 p.m., INR 038)

Problem 1 (PURE RANDOMNESS AND BIASED COINS)

Let $X_1, X_2, \ldots, X_n$ denote the outcomes of independent flips of a biased coin. Thus, $Pr\{X_i = 1\} = p$, $Pr\{X_i = 0\} = 1 - p$ where p is unknown. We wish to obtain a sequence $Z_1, Z_2, \ldots, Z_K$ of *fair* coin flips from $X_1, X_2, \ldots, X_n$. Towards this end, let $f : \mathcal{X}^n \to [0,1]^*$ (where $[0,1]^* = [\Lambda, 0, 1, 00, 01, \ldots]$ is the set of all finite-length binary sequences, where Λ is the null string) be a mapping $f(X_1, X_2, \ldots, X_n) = (Z_1, Z_2, \ldots, Z_K)$, where $Z_i \sim$ Bernoulli $(\frac{1}{2})$, and K may depend on $(X_1, X_2, \ldots, X_n)$. In order that the sequence $Z_1, Z_2, \ldots$ appear to be fair coin flips, the map f from biased coin flips to fair coin flips must have the property that all 2^k sequences $Z_1, Z_2, \ldots, Z_k$ of a given length k have equal probability (possibly 0), for $k = 1, 2, \ldots$. For example, for $n = 2$ the map $f(01) = 0$, $f(10) = 1$, $f(00) = f(11) = \Lambda$ has the property that $Pr\{Z_1 = 1|K = 1\} = Pr\{Z_1 = 0|K = 1\} = \frac{1}{2}$. Give reasons for the following inequalities:

$$\begin{aligned}
nH(p) &\overset{(a)}{=} H(X_1, X_2, \ldots, X_n) \\
&\overset{(b)}{\geq} H(Z_1, Z_2, \ldots, Z_K, K) \\
&\overset{(c)}{=} H(K) + H(Z_1, Z_2, \ldots, Z_K|K) \\
&\overset{(d)}{=} H(K) + \mathbb{E}[K] \\
&\overset{(e)}{\geq} \mathbb{E}[K],
\end{aligned}$$

where $\mathbb{E}$ is the expectation operator. Thus, no more than $nH(p)$ fair coin tosses can be derived from $(X_1, X_2, \ldots, X_n)$, on the average. Exhibit a good map f on sequences of length 4.

Problem 2 (INEQUALITIES)

Let X, Y, and Z be joint random variables.

(a) Prove the following inequalities and find conditions for equality.

1. $H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X)$.
2. $I(X;Z|Y) \geq I(Z;Y|X) - I(Z;Y) + I(X;Z)$.

(b) Give examples of X, Y, and Z such that

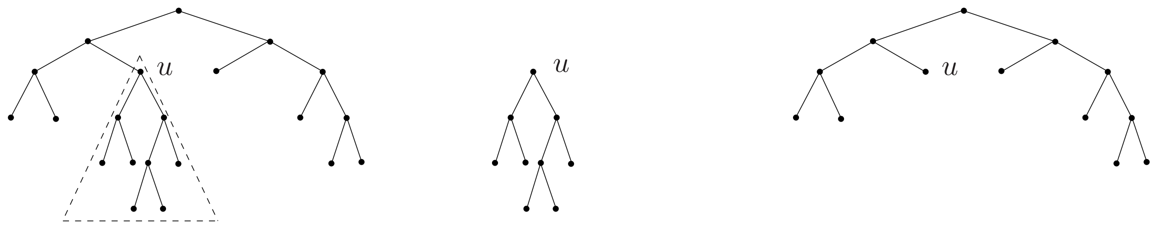
1. $I(X;Y|Z) < I(X;Y)$.
2. $I(X;Y|Z) > I(X;Y)$.

Problem 3 (HUFFMAN SUB-TREE)

Let $\mathcal{S}$ be a source with alphabet $\{x_1, \dots, x_n\}$, with associated probabilities $\mathcal{P} = (p_1, \dots, p_n)$. We compress this source using a binary Huffman code, were a source symbol x_i is associated with a codeword $c_i(x_i)$ of length ℓ_i. Denote the corresponding binary tree by $\mathcal{T}$.

(a) Write expressions for the $L(\mathcal{P})$, average length of the code, and $H(\mathcal{P})$, the entropy of the source, in terms of ℓ_i's and p_i's.

Denote the corresponding binary tree by $\mathcal{T}$. Let u be an intermediate node in the tree of distance ℓ from the root, and denote by $\mathcal{T}_u$ the sub-tree below u, and by $\mathcal{S}_u$ the set of source symbols located on the leaves of this sub-tree, as shown in Fig. 3. Assume $\mathcal{S}_u = \{x_k + 1, \dots, x_n\}$. Also let $\mathcal{T}^u$ be the same tree unless the sub-tree below u is merged in a node u, with probability $q = \sum_{i=k+1}^{n} p_i$.



(a) $\mathcal{T}$: A Huffman tree (b) $\mathcal{T}_u$: the sub-tree below u (c) $\mathcal{T}^u$: the Huffman tree merged at u

(b) Argue that Huffman tree $\mathcal{T}^u$ is a valid Huffman code tree for the source $\mathcal{S}^u = \{x_1, \dots, x_k, u\}$, with probability distribution $\mathcal{P}^u = (p_1, \dots, p_k, q)$.

(c) Express the $L(\mathcal{P}^u)$ and $H(\mathcal{P}^u)$, the average length and entropy of the the source $\mathcal{S}^u$, in terms of ℓ_i's, p_i's, ℓ, and q.

(d) Argue that the sub-tree $\mathcal{T}_u$ is a valid Huffman code tree for the source $\mathcal{S}_u$, with probability distribution $\mathcal{P}_u = (\frac{p_{k+1}}{q}, \frac{p_{k+2}}{q}, \dots, \frac{p_n}{q})$, where $q = \sum_{i=k+1}^{n} p_i$.

(e) Express $L(\mathcal{P}_u)$ and $H(\mathcal{P}_u)$, the average length and entropy of the the source $\mathcal{S}_u$, in terms of ℓ_i's, p_i's, ℓ, and q.

(f) How can we relate the entropy of the sources $\mathcal{S}_u$ and $\mathcal{S}^u$ to the entropy of the original source, $\mathcal{S}$? Form a similar expression to relate the average lengths.

Problem 4 (SUFFICIENT STATISTICS)

Suppose that we have a family of probability mass functions $\{f_\theta(x)\}$ indexed by θ, and let X be a sample from a distribution in this family. Let $T(X)$ be any statistic (e.g. sample mean or sample variance is a possible statistic.)

(a) Show that

$$I(\theta; T(X)) \leq I(\theta; X)$$

for any distribution on θ.
A statistic $T(X)$ is called sufficient if equality holds for any distribution on θ, or equivalently if $\theta \to T(X) \to X$ forms a Markov chain for all distributions on θ.

(b) Let $X_1, X_2, \cdots, X_n$, $X_i \in \{0, 1\}$, be an independent and identically distributed (i.i.d.) sequence of coin tosses of a coin with an unknown parameter $p = pr(X_i = 1)$. Show that the number of 1's ($\sum_{i=1}^{n} X_i$) is a sufficient statistic for p.