ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

**Handout 18** Information Theory and Coding

Midterm Solutions Nov. 13, 2012

PROBLEM 1.

(a) It suffices to check that the lengths $\ell(u^n) = \left\lceil \log_2(n+1) + \log_2 \binom{n}{k} \right\rceil$ (with $k$ being the number of 1's in $u^n$ satisfy Kraft's inequality:

$$\begin{aligned}
\sum_{u^n} 2^{-\ell(u^n)} &= \sum_{k=0}^{n} \sum_{\substack{u^n:\, u^n \\ \text{has } k \text{ 1's}}} 2^{-\ell(u^n)} \\
&\le \sum_{k=0}^{n} \sum_{\substack{u^n:\, u^n \\ \text{has } k \text{ 1's}}} \frac{1}{n+1}\binom{n}{k}^{-1} && \text{since } \ell(u^n) \ge \log(n+1) + \log\binom{n}{k} \\
&= \sum_{k=0}^{n} \frac{1}{n+1} && \text{since the number of } u^n \text{ with } k \text{ 1's is } \binom{n}{k} \\
&= 1,
\end{aligned}$$

and we conclude that a prefix-free code with the desired lengths exists.

The form of $\ell(u^n)$ suggests the following way to implement $\mathcal{C}$ (ignoring integer constraints): given $u^n$ first describe the number of 1's in it, $k$, by using $\log_2(n+1)$ bits (note that $k$ takes on only $n+1$ possible values, 0, 1, ..., $n$.) Next descibe the sequence $u^n$ by giving its index among the $\binom{n}{k}$ sequences with $k$ 1's and $n-k$ 0's.

(b&c) Taking the hint $1 = \sum_{i=0}^{n} \binom{n}{i} z^i (1-z)^{n-i} \ge \binom{n}{k} z^k (1-z)^{n-k}$. which is equivalent to what is to be shown in (b). Choosing $z = k/n$ gives $\binom{n}{k} 2^{nh_2(k/n)} \le 1$, which is what is to be shown in (c).

(d) Since $U^n$ is random the number of 1's in it, $K$, is also random with

$$E[K] = E\big[\sum_i U_i\big] = \sum_i E[U_i] = np.$$

We also have

$$\begin{aligned}
\text{length}(\mathcal{C}_n(U^n)) &< 1 + \log(n+1) + \log\binom{n}{K} \\
&\le 1 + \log(n+1) + nh_2(K/n) \\
E\big[\text{length}(\mathcal{C}_n(U^n))\big] &< 1 + \log(n+1) + nE[h_2(K/n)] \\
&\le 1 + \log(n+1) + nh_2(E[K]/n) && \text{concavity of } h_2 \\
&= 1 + \log(n+1) + nh_2(p),
\end{aligned}$$

as was to be shown.

Note that we have shown that the code we construced compresses any i.i.d. binary source to close to its entropy (asymptotically, exactly to its entropy), and so it is universal for this class of sources.

Problem 2.

(a) An element $x^n$ of $T_2$ has at most $n(1-q)(1+\epsilon)$ 0's and at most $nq(1+\epsilon)$ 1's. Consequently, $c(x^n)$ is at most $[(1-q)nc(0)+qnc(1)](1+\epsilon)$. (A lower bound of $[(1-q)nc(0)+qnc(1)](1-\epsilon)$ can be derived in the same way.)

(b) From the properties of typical sets, for large $n$ the cardinality of $T_2$ is at least $(1-\epsilon)2^{n(1-\epsilon)h_2(q)}$, whereas the cardinality of typical source sequences $T(m)$ is at most $2^{m(1+\epsilon)H(U)}$. Consequently, the condition guarantees that there are at least as many sequences in $T_2$ as in $T(m)$ and thus that each typical source sequence can be assigned a distinct representative in $T_2$.

(c) By construction, each typical source sequence of length $m$ is assigned a binary sequence of cost at most $n\big[(1-q)c(0)+qc(q)\big](1+\epsilon)$, which yields the claimed cost per source letter.

(d) If we choose

$$n = n(m,\epsilon) := \left\lceil m\frac{(1+\epsilon)H(U)}{(1-\epsilon)h_2(q)}\right\rceil$$

we are assured of satisfying the sufficient condition. Note that $\lim_{\epsilon\to 0}\lim_{m\to\infty} n(m,\epsilon)/m = H(U)/h_2(q)$. Thus, given $\delta$, if $\epsilon$ is chosen sufficiently small and $m$ sufficiently large we can ensure that

(i) $\dfrac{n}{m} \leq \dfrac{H(U)}{h_2(q)}(1+\delta/3)$,

(ii) $(1+\epsilon)$ that appears as a factor in the upper bound to cost is smaller than $(1+\delta)/(1+\delta/3)$.

(iii) $\Pr(U^m \in T(m,\epsilon,p_U)) \geq 1-\delta$ (so that the probability that $U^m$ has no representative has negligible probability.)

(e) We can handle to non-typical $u^m$'s by prefixing the binary sequences assigned to typical $u^m$'s by 1, (this changes the cost per letter in a neglible way for large $m$), and then assigning to non-typical $u^m$ binary sequences that start with 0, followed by a unique $m\log_2|\mathcal{U}|$ bit representation. The cost of the binary representatives for such non-typical sequences is at most $(m+1)\log_2|\mathcal{U}|\max\{c(0),c(1)\}$ and thus the cost per letter is bounded. Since the set of non-typical $u^m$'s has negligible probability, they make a neglibible contribution to the expected cost.

(f) To find the cost per information bit, we further normalize the expression in (d) by $H(U)$, allowing us to conclude that for any $\delta$ and any $q$ there is a method of representing the source with a cost of

$$\frac{(1-q)c(0)+qc(1)}{h_2(q)}(1+\delta)$$

per information bit. Since we can freely choose $q$, we can choose it to minimize this expression, yielding the cheapest possible representation within the strategies we have considered.

A finer argument also yields that no strategy (fixed-to-fixed or not) can represent information at a lower cost, so the bound we derived is in general the best possible. Unequal cost bits are not uncommon: think of the Morse code where the time it takes to send a dash is longer than a dot.

Problem 3.

(a) Note that $X_i = x$ if and only if $c_i(W) = x$. By the definition of $p_i$, there are exactly $Mp_i(x)$ messages $m$ for which $c_i(m) = x$. Since $W$ is uniformly distributed on $\{1, \dots, M\}$ the probability that $c_i(W) = x$ is $Mp_i(x)/M = p_i(x)$.

(b) We know, by the data processing theorem, that $I(W; Y^n) \le I(X^n; Y^n)$. We also know, from class, that when $X^n$ and $Y^n$ are the input and output of a memoryless channel, $I(X^n; Y^n) \le \sum_i I(X_i; Y_i)$. Since $I(X_i; Y_i)$ is the mutual information between the input and output of the channel and since $X_i$ has distribution $p_i$, the value of $I(X_i; Y_i)$ is $I_i$ by the definition of $I_i$. Thus,

$$I(W; Y^n) \le I(X^n; Y^n) \le \sum_i I(X_i; Y_i) = \sum_{i=1}^{n} I_i.$$

as was to be shown.

(c) Denoting by $I(p)$ the mutual information between the input and output of the channel with the input has distribution $p$, we have $I_i = I(p_i)$ and $\bar{I} = I(\frac{1}{n} \sum_i p_i)$. From class, we know that $I(p)$ is a concave function of $p$, and thus

$$\bar{I} \ge \frac{1}{n} \sum_i I(p_i) = \frac{1}{n} \sum_i I_i.$$

(d) By (c), the mutual information achieved by the engineer's code is upper bounded by

$$\frac{1}{n} I(W; Y^n) \le I(0.7, 0.3) := I(X; Y)|_{(p_X(0), p_X(1)) = (0.7, 0.3)} < C$$

as the capacity of the binary symmetric channel is achieved by the distribution that makes the inputs 0 and 1 equally likely. The engineer's scheme can at best achieve the rate $I(0.7, 0.3)$.

(e) We have already observed that $I_i = I(p_i)$. Noting that

$p_1$ is determined by $c_1(1), \dots, c_1(M)$,
$p_2$ is determined by $c_2(1), \dots, c_2(M)$,
...
$p_n$ is determined by $c_n(1), \dots, c_n(M)$,

we see that $p_1$, $p_2$, ..., $p_n$ are independent of each other and have identical distribution. Consequently $I_i$ are i.i.d.

Futhermore, for any $x \in \mathcal{X}$, $p_i(x) = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\{c_i(m) = x\}$, and thus

$$E[p_i(x)] = \frac{1}{M} \sum_{m=1}^{M} \Pr\{c_i(m) = x\} = p(x).$$

Consequently, $E[I_i] = E[I(p_i)] \le I(E[p_i]) = I(p)$, as was to be shown. (The inequality is due to, again, the concavity of $I(p)$ as a function of $p$.)

What is shown here, combined with the law of large numbers, shows that if that a code constructed randomly by choosing each letter of each codeword independently according to a distribution $p$, then with very high probability the code will have $\frac{1}{n}I(W;Y^n) \leq I(p)$. This conclusion can then be use to show the existence of codes that keep a bad receiver (e.g., an eavesdropper) ignorant of the transmitted data while reliably sending data to a good receiver.