Essential Coding Theory — Madhu Sudan
6.896
Extended Due Date: Wednesday, October 2, 2002

Problem Set 2

Instructions: See PS 1.

Problems

1. Prove the noiseless coding theorem, and its converse. (But don't turn in.)

2. Consider a Markovian source of bits, where the source consists of a 6-cycle with three successive vertices outputing 0, and three successive vertices outputting 1, with the probability of either going left (or right) from any vertex is exactly 1/2. Compute the rate of this source. (I expect an ab initio argument. Hopefully this will motivate you to look up Shannon's general method for computing the rate of a Markovian source.)

   Basic idea of ad-hoc analysis: Compress state diagram to a minimum and then make some basic observations to compress the source information.

   The state diagram can be compressed into an equivalent one on four states, say E_0, M_0, M_1, and E_1, (E for End, and M for Middle) where states E_i and M_i generate bit i; and from state E_i the chain jumps deterministically to state M_i at the next step; and state M_i jumps to state E_i with probability half (Type-1 move) and to state M_{1-i} with probability half (Type-2 move).

   To describe a string generates by the source, notice it suffices to give the start state (one of four possibilities) and the sequence of moves made from states M_i. After ℓ visits to the states $\{M_0, M_1\}$, the expected number of Type-1 moves is $\ell/2$ and number of Type-2 moves is $\ell/2$. Thus the expected length of the output string after ℓ visits to the M states is $2 \times (\ell/2) + \ell/2$. Thus a sequence of ℓ bits (+ two for the initial state) suffices to describe $3\ell/2$ output bits (in expectation, or with high probability). Inverting this gives a compression rate of $2/3$.

   It is also clear one can't do better. Any way to describe the output string, describes the sequence of Type-1/Type-2 moves made from the M states. W.h.p., $2/3$rds of the time is spent in such states. and this sequence of moves is a random, unbiased, independent sequence of bits.

   On systematic analyses: Shannon's original paper shows (claims?) that the rate of such a source is the limit of the following quantity: Let D_n denote the distribution of n bit strings generated by the source after starting at some fixed state and n time steps. Let $H(D_n)$ denote the entropy of this distribution. Let R_n denote $H(D_n)/n$. Then the rate of this source, R, equals $\lim_{n\to\infty}\{R_n\}$. It is possible to write a set of recurrences that allow one to describe the entropy of D_n in terms of the entropy of D_{n-1} and this gives a systematic way to show this bound.

3. Consider a binary channel whose input/output alphabet is $\{0,1\}$, where a 0 is transmitted faithfully as a 0 (with probability 1), but a 1 is transmitted as a 0 with probability $\frac{1}{2}$ and a 1 with probability 1/2. Compute the capacity of this channel. (You should prove this from scratch using only simple probabilistic facts already stated/used in class - not by referring to tools gleaned from other courses in information theory. For partial credit, you may just prove a lower bound on the capacity. The higher your bound, the more the credit.)

   Lets pick the encoding function $E : \{0,1\}^k \to \{0,1\}^n$ at random with $E(x)$ being independent of $E(x')$. However, $E(x)$ is not uniformly chosen. Instead, we will pick $E(x)$ uniformly from $W_{p,n}$, i.e., strings of weight pn in $\{0,1\}^n$.

   We say x is consistent with a received string r if $r_i = 0$ whenever $E(x)_i = 0$. The decoding algorithm works as follows: (1) If a received vector $r \in \{0,1\}^n$ does not have weight in the interval $[(p/2-\epsilon)n, (p/2+\epsilon)n]$ then declare an error. (2) If there exist two distinct vectors x, x' that are consistent with r, declare an error. (3) Else, compute an x that is consistent with r and output x.

   For the analysis, first note that the message x sent over the channel is always consistent with the received vector, so the Steps (2) and (3) are exhaustive. The probability of error declared in Step (1) is exponentially small, so we can just about ignore it. Error in Step (2) occurs if some x' other than the transmitted message is consistent with a received vector r. For fixed x', if r has weight w, this probability is exactly $\binom{n-w}{pn-w}/\binom{n}{pn}$, which happens to equal $\binom{pn}{w}/\binom{n}{w}$. (Go figure this one out!) Using $w \approx pn/2$ and $\binom{n}{\alpha n} \approx 2^{H(\alpha)n}$, we get that this event happens with probability approximately $2^{(p-H(p/2))n}$. The probability that such an x' exists is thus $2^{k-(H(p/2)-p)n}$. Thus $k/n < H(p/2) - p$ would lead to exponentially small error in Step 2 also. Thus a rate of $H(p/2) - p$ is achievable. Some differential calculus yields that this expression is maximized at $p = 2/5$ giving a rate of $H(1/5) - 2/5$.

   To show that this is the upper bound fix an encoding and decoding pair (E, D). Suppose E maps at least $1/(n+1)$ fraction of the messages to vectors of weight pn (such a p must exist!). Let this subset of messages be M with size K. Consider picking a random message from this space and encoding and transmitting it. With all but exponentially small probability the received vector has weight in $[(p/2-\epsilon)n, (p/2+\epsilon)n]$. Furthermore, by symmetry, the probability that any specific vector is the one received given a fixed message from M, is exactly 2^{-pn}. Thus the probability of successful decoding can be calculated exactly to be $K^{-1}2^{-pn}\binom{n}{pn/2}$ which is roughly the reciprocal of the quantity we got in the forward direction. So K better be at most $2^{H(p/2)-p}n$ giving that the capacity computed above is tight.

4. If there is a constructive solution to Shannon's noisy coding theorem with E being a linear map, then show that there is a constructive solution to Shannon's noiseless coding theorem in the case where the source produces a sequence of independent bits of bias p.

   Clarifications:

   (a) The encoding and decoding functions used in the noiseless theorem should be polynomial time computable, if the corresponding functions are polynomial time computable in the noisy theorem.

   (b) The compression rate in the noiseless coding theorem should be arbitrarily close to $H(p)$, assuming the rate of the encoding function in the coding theorem can be made arbitrarily close to $1 - H(p)$.

Some further caveats. I should have stated explicitly that we would allow some small probability of decoding error, (Either that, or I need to use non-uniformity in the choice of the encoding/decoding algorithm.)

Let the encoding function correspond to a linear code with $n \times (n-k)$ parity check matrix H. Given a vector η from the p-biased source, its compression $E'(\eta)$ will be the vector ηH. The decompression algorithm $D'(y)$ is computed as follows: First compute (by simple linear algebra) a random vector r such that $rH = y$. Now it decode r to obtain $c = D(r)$ and let $\eta' = r - c$ where c is a codeword. To analyze the probability that $D'(E'(\eta)) \neq \eta$, note that the $c' = r - \eta$ is a random codeword under E, and η is a random error vector under the binary symmetric channel. Thus the probability that $D(c' + \eta) \neq c'$ is negligibly small. If this event does not occur, we have $D(r) = r - \eta$ and then $\eta' = \eta$ as desired.

5. Given codes C_1 and C_2 with encoding functions $E_1 : \{0,1\}^{k_1} \to \{0,1\}^{n_1}$ and $E_2 : \{0,1\}^{k_2} \to \{0,1\}^{n_2}$ let $E_1 \otimes E_2 : \{0,1\}^{k_1 \times k_2} \to \{0,1\}^{n_1 \times n_2}$ be the encoding function obtained as follows: View a message $\mathbf{m}$ as a $k_1 \times k_2$ matrix. Encode the columns of $\mathbf{m}$ individually using the function E_1 to get an $n_1 \times k_2$ matrix $\mathbf{m}'$. Now encode the rows of $\mathbf{m}'$ individually using E_2 to get an $n_1 \times n_2$ matrix that is the final encoding under $E_1 \otimes E_2$ of $\mathbf{m}$. Let $C_1 \otimes C_2$ be the code associated with $E_1 \otimes E_2$.

   For $i \geq 3$, let H_i denote the $[2^i - 1, 2^i - i - 1, 3]_2$-Hamming code. Let $C_i = H_i \otimes C_{i-1}$ with $C_3 = H_3$ be a new family of codes.

   (a) Give a lower bound on the relative minimum distance of C_i. Does it go to zero as $i \to \infty$?

   As noted in class, the product of two codes of minimum distance d_1 and d_2 respectively is at least $d_1 d_2$, and there exist codes for which the product has mininum distance exactly $d_1 d_2$ (for all linear codes, in fact).
   So C_i has mininum distance 3^{i-2}. Its block length is roughly $\prod_{j=3}^{i} 2^j = 2^{\Theta(i^2)}$. Thus the relative minimum distance of C_i is $2^{O(i)-\Theta(i^2)}$ which goes to zero as $i \to \infty$.

   (b) Give a lower bound on the rate of C_i. Does it go to zero as $i \to \infty$?

   The rate of $H_i = 1 - \frac{i}{2^i-1} \geq 1 - \frac{i}{2^{i-1}}$. The rate of C_i is thus at least $\prod_{j=3}^{i}(1 - \frac{j}{2^{j-1}})$. Using the formal series $\frac{1}{1-x} = \sum_{j=0}^{\infty} x^j$ and $\frac{1}{(1-x)^2} = (\sum_{j=0}^{\infty} x^j)^2 = \sum_{j=0}^{\infty}(j+1)x^j$, we get $\sum_{j=1}^{\infty} \frac{j}{2^{j-1}} = 4 < \infty$. we get $\sum_{j=5}^{\infty} \frac{j}{2^{j-1}} = 3/4$. So we get the rate of C_i is at least $\prod_{j=3}^{i}(1 - \frac{j}{2^{j-1}}) \leq (1 - 3/4)(1 - 4/8)(1 - \sum_{j=5}^{\infty} \frac{j}{2^{j-1}}) = \frac{1}{32}$.

   (c) Consider the following simple decoding algorithm for C_i: Decode the rows of the rec'd vector recursively using the decoding algorithm for C_{i-1}. Then decode each column according to the Hamming decoding algorithm. Let p_i denote the probability of decoding error of this algorithm on the Binary Symmmetric Channel with parameter p. Show that there exists a $p > 0$ such that $p_i \to 0$ as $i \to \infty$. (Hint: First show that $p_i \leq 4^i p_{i-1}^2$.)

   First note that we have a decoding error when using code C_i only if two of the rows of the rec'd vector are decoded incorrectly. For any fixed row, the probability of a decoding error is p_{i-1}. The probability of decoding two fixed distinct rows

incorrectly is thus p_{i-1}^2. Since the code has $2^i - 1$ rows, the probability there exist a pair of rows that are decoded incorrectly is thus at most $4^i p_{i-1}^2$ yielding the bound in the hint.
We'll prove by induction that $p_i \le 2^{-2^{i/2}}$. The base cases (for small i) follow if p is small enough. For induction, we have $p_i \le 4^i 2^{-2\cdot 2^{(i-1)/2}} = 2^{-(\sqrt{2}2^{i/2}-2i)} \le 2^{-2^{i/2}}$ where the last inequality holds provided i is large enough so as to satisfy $(\sqrt{2}-1)2^{i/2} \ge 2i$. Clearly $p_i \to 0$ as $i \to \infty$. (In fact, if we express the error probability as a function of the block length, we have the error shrinks as $\exp(-2^{\sqrt{\log n}}$ for codes of length n, which is not exponential in n, but better than $1/n^c$ for any c!)