TUTORIAL XIII

Problem 1 (True or false)**.** For each one of these statements, say whether it is true or false and provide a brief justification.

1. Define the distribution $P_X = (1/6, 1/6, 1/6, 3/6)$. We have $H(X) = \log_2 5$. False, $H(X) \leq \log |\mathcal{X}|$
2. Define the channel W with binary input and output given by $W(0|0) = 1/3, W(1|0) = 2/3, W(0|1) = 2/5, W(1|1) = 3/5$. The capacity of this channel is 0. False, nontrivial channel.
3. Let C be a randomly chosen binary code with blocklength n and dimension $n/2$, i.e., a uniformly distributed subset of $\{0,1\}^n$ of size $2^{n/2}$. Then, with probability going to 1 as $n \to \infty$, C is not a linear code. True, $\mathbf{P}\{C \text{ is linear}\} \leq \frac{2^{n^2/2}}{\binom{2^n}{2^{n/2}}} \to 0$.
4. Consider the distribution $P_X = (1/2, 1/6, 1/6, 1/6)$. The code with the shortest expected length for this source has expected length exactly $H(X)$. False, $H(X) = \frac{1}{2} + \frac{3}{6}\log_2 6$ is not rational.
5. Let $X_1, \ldots, X_n$ be iid boolean random variables with distribution $P_{X_1}(0) = 1/4$ and $P_{X_1}(1) = 3/4$. Let $(x_1, \ldots, x_n) \in \{0,1\}^n$ be such that $|\{i \in \{1, \ldots, n\} : x_i = 0\}| = n/2$. Then, for large enough n, $(x_1, \ldots, x_n)$ is $\frac{1}{100}$-typical, i.e., $2^{-n(H(X_1)+\frac{1}{100})} \leq P_{X_1 \ldots X_n}(x_1, \ldots, x_n) \leq 2^{-n(H(X_1)-\frac{1}{100})}$. False, $H(X_1) = 2 - 3/4 \log 3$ and $P(x_1 \ldots x_n) = 2^{-(2-\log_2 3)n}$

Problem 2 (Basic entropy computation)**.** Suppose we have a competition between 3 teams named a, b and c. Each team plays once against the two other teams so that there are 3 games in total. A game between 2 teams, say a and b, has only two possible outcomes: either a wins or b wins. We denote the outcome of the games by X_{ab}, X_{ac}, X_{bc}. For example $X_{ab} \in \{a, b\}$ denoting the winning team. We take a simple model where X_{ab}, X_{ac} and X_{bc} are independent and uniformly distributed. We then define the scores S_a, S_b, S_c for each team to be the number of games won by the team. Let $W \in \{a, b, c, d\}$ denote the team that has the largest score and d if there is a tie.

Compute $H(X_{ab}, X_{ac}, X_{bc}), H(S_a, S_b, S_c | X_{ab}, X_{ac}, X_{bc}), H(S_a, S_b, S_c), H(W|S_a, S_b, S_c)$ and $H(W)$. $3, 0, 11/4, 0, 2$

Problem 3 (Repetition code)**.** Let $C_k^{(r)}$ be a binary repetition code whose encoding function repeats each bit of the message r times. More precisely, for a bitstring $m_1 \ldots m_k \in \{0,1\}^k$, let $C_k^{(r)}(m_1 \ldots m_k) = m_1^{(r)} \ldots m_k^{(r)} \in \{0,1\}^{rk}$, where $m^{(r)}$ denotes the concatenation of r copies of the bit m.

1. Show that $C_k^{(r)}$ is a linear code with minimum distance r. In other words, it is a $[rk, k, r]_2$ code.
2. Write a generator matrix and a parity check matrix for $C_k^{(r)}$.
3. Recall that $\text{BSC}_f(b|b) = 1 - f$ and $\text{BSC}_f(1-b|b) = f$ for any $b \in \{0,1\}$. We would like to know if it is a good idea to use a code $C_k^{(r)}$ to achieve reliable communication close to the capacity of the channel $\text{BSC}_{0.25}$. What is the capacity of the channel $\text{BSC}_{0.25}$?
4. Given that $\frac{1}{9} \approx 0.111$ and $1 - H_2(0.25) \approx 0.189$, let us choose $r = 9$ to code at a rate not too far from the capacity. If we use the code $C_k^{(9)}$ to transmit k bits over $9k$ copies of $\text{BSC}_{0.25}$, can we make the error probability for decoding go to 0 as $k \to \infty$? No, with probability 0.25^9 one can change from a codeword to a different codeword so no decoder can have a better success probability.

Problem 4 (Constructing good codes)**.** The objective of this problem is to explicitly construct a family of binary linear codes with dimension $k = \Omega(n)$ and minimum distance $d = \Omega(n)$.

1. We will define a family of codes with blocklength $2k$ and dimension k. Recall that we can view the set $\{0,1\}^k$ as a field $\mathbb{F}_{2^k}$ (the only thing needed for this problem is that it is a field). More formally, we assume that $\sigma : \mathbb{F}_2^k \to \mathbb{F}_{2^k}$ is a bijection and satisfies the properties $\sigma(0) = 0$, $\sigma(x+y) = \sigma(x)+\sigma(y)$ for any $x, y \in \mathbb{F}_2^k$ and also $\sigma^{-1}(u+v) = \sigma^{-1}(u) + \sigma^{-1}(v)$ for $u, v \in \mathbb{F}_{2^k}$. For every $\alpha \in \mathbb{F}_{2^k}$ nonzero, let $C_\alpha : \{0,1\}^k \to \{0,1\}^{2k}$ be defined by $C_\alpha(x) = (x, \sigma^{-1}(\alpha \cdot \sigma(x)))$. Here $\cdot$ denotes the multiplication in the field $\mathbb{F}_{2^k}$.

   (a) Show that for any α, C_α is a linear code. For $\alpha = 1$ (the unit for the field $\mathbb{F}_{2^k}$), what is the minimum distance of C_1?

   (b) Show that for $\alpha \neq \beta$, $C_\alpha \cap C_\beta = \{0\}$.

   (c) Show that the fraction of codes C_α with minimum distance $\leq d-1$ is at most $\frac{\sum_{i=1}^{d-1}\binom{2k}{i}}{2^k-1}$. Recall that for large enough k, $\sum_{i=0}^{d-1}\binom{2k}{i} \leq 2^{2kH_2(\frac{d}{2k})}$. Let $\epsilon > 0$ and $d = H_2^{-1}(\frac{1}{2} - \epsilon)2k$. Show that the fraction of codes with minimum distance $\geq d$ is at least $1 - 2^{-\epsilon k}$.

$$|\{\alpha \in \mathbb{F}_{2^k} - \{0\} : d(C_\alpha) \leq d-1\}| \leq |\{x \in \{0,1\}^{2k} : |x| \leq d-1\}| = \sum_{i=1}^{d-1}\binom{2k}{i}. \tag{1}$$

   Then

$$\frac{\sum_{i=0}^{d-1}\binom{2k}{i} - 1}{2^k - 1} \leq \frac{\sum_{i=0}^{d-1}\binom{2k}{i}}{2^k-1} \leq \frac{2^{2k(1/2-\epsilon)}}{2^k} = 2^{-2k\epsilon}. \tag{2}$$

2. The problem is this family is that we do not know which value of α leads to a good code. Let RS be a Reed Solomon $[2^k - 1, 2^{k-1}, 2^{k-1}]_{2^k}$ code.

   (a) Give a generator matrix for the code RS.

   (b) Consider the concatenation of the code RS and use as inner codes the codes C_α, i.e., the block labeled α is encoded using the code C_α. The resulting code is a binary code. What is the blocklength and the dimension of the resulting code? Give a lower bound on the minimum distance that is linear in the blocklength. Blocklength: $(2^k - 1)2k$ bits, dimension: $2^{k-1}k$. Minimum distance: Encoding a nonzero message, there are at least 2^{k-1} symbols that are nonzero. Take $\epsilon = 1/4$, then for large enough k, at least $9/10$ of the α's have a minimum distance $\geq H_2^{-1}(1/4) \cdot 2k$. So out of the 2^{k-1}, at least $4/52^{k-1}$ give an encoded block of weight at least $H_2^{-1}(1/4)2k$ which gives an overall minimum distance of $4/52^{k-1}H_2^{-1}(1/4)2k$, which is linear in k.