Tutorial XIV: Revision for Final

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

1. There are at most 2^{nk} binary linear codes of blocklength n and dimension k.
2. 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.
3. 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 $H(X)$.
4. Let W be a channel with binary input and output such that $W(0|0) \neq W(0|1)$, i.e., the output distributions are different for different inputs. The capacity of this channel is > 0.
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., $2^{-n(H(X_1)+\frac{1}{100})} \leq P_{X_1 \ldots X_n}(x_1, \ldots, x_n) \leq 2^{-n(H(X_1)-\frac{1}{100})}$.
6. Let $G = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$. The binary code whose generator matrix is G has a minimum distance of 4.
7. The code $C = \{0000, 0011, 1111\}$ can detect any error on two bits.
8. The code over $\mathbb{F}_5$ with generator matrix $G = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}$ has a minimum distance of 3 and among all codes over $\mathbb{F}_5$ with the same blocklength and dimension it has the largest possible minimum distance.
9. For any random variable $X \in \mathcal{X}$, there exists an $x \in \mathcal{X}$ such that $P_X(x) \leq 2^{-H(X)}$.

Problem 2 (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 $\mathrm{BSC}_f(b|b) = 1 - f$ and $\mathrm{BSC}_f(1-b|b) = f$ 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 $\mathrm{BSC}_{0.25}$. What is the capacity of the channel $\mathrm{BSC}_{0.25}$?
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 $\mathrm{BSC}_{0.25}$, can we make the error probability for decoding go to 0 as $k \to \infty$?

Problem 3 (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 $d = H_2^{-1}(\frac{1}{2} - \frac{\epsilon}{2})2k$. Show that the fraction of codes with minimum distance $\geq d$ is at least $1 - 2^{-\epsilon k}$.

2. The problem in 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.

Problem 4. Given two channels $W^1_{Y_1|X_1}$ and $W^2_{Y_2|X_2}$ with input spaces $\mathcal{X}_1, \mathcal{X}_2$ and outputs spaces $\mathcal{Y}_1, \mathcal{Y}_2$. Consider the channel W^{12} defined on input space $\mathcal{X}_1 \times \mathcal{X}_2$ and output space $\mathcal{Y}_1 \times \mathcal{Y}_2$ and $W^{12}_{Y_1Y_2|X_1X_2}(y_1y_2|x_1x_2) = W^1_{Y_1|X_1}(y_1|x_1) \cdot W^2_{Y_2|X_2}(y_2|x_2)$. Compute $\mathrm{C}(W^{12}) = \max_{P_{X_1X_2}} I(X_1X_2 : Y_1Y_2)$ (where Y_1Y_2 is the output of W^{12} when the input is X_1X_2) as a function $\mathrm{C}(W^1)$ and $\mathrm{C}(W^2)$.

Problem 5. Let C be an $[n,k]_2$ linear code with w_j denoting the number of codewords of C of Hamming weight j for $0 \leq j \leq n$. Define the polynomial $f(X) = \sum_{j=0}^{n} w_j X^j$.

1. What is the value of w_0 and $\sum_{j=0}^{n} w_j$?

2. Suppose C is used for the transmission over n copies of the binary symmetric channel with flip probability $p < \frac{1}{2}$ and that we use a maximum likelihood decoder, i.e., given $y \in \{0,1\}^n$, the decoder outputs $c \in C$ such that the Hamming distance between y and c is minimized. Show that for any transmitted codeword, the probability of an incorrect decoding is at most $f(\xi) - 1$ with $\xi = \sqrt{4p(1-p)}$.