Today

- Asymptotically good codes.
- Random/Greedy codes.
- Some impossibility results.

Rate and Relative Distance

Recall the four integer parameters

- (Block) Length of code n
- Message length of code k
- Minimum Distance of code d
- Alphabet size q

Code with above parameters referred to as $(n, k, d)_q$ code. If code is linear it is an $[n, k, d]_q$ code.

(Deviation from standard coding non-linear codes are referred to by number of codewords. so a linear $[n, k, d]_q$ with the all zeroes word

deleted would be an $(n, q^k - 1, d)_q$ code, while we would have it as an $(n, k - \epsilon, d)_q$ code.)

Today will focus on the normalizations:

- Rate $R \stackrel{\text{def}}{=} k/n$.
- Relative Distance $\delta \stackrel{\text{def}}{=} d/n$.

Main question(s): How does R vary as function of δ, and how does this variation depend on q?

Impossibility result 1: Singleton Bound

Note: Singleton is a person's name! Not related to proof technique. Should be called "Projection bound".

Main result: $R + \delta \leq 1$.

More precisely, for any $(n, k, d)_q$ code, $k + d \leq n + 1$.

Proof: Take an $(n, k, d)_q$ code and project on to $k - 1$ coordinates. Two codewords must project to same sequence (PHP). Thus these two codewords differ on at most $n - (k - 1)$ coordinates. Thus $d \leq n - k + 1$.

Impossibility result 2: Hamming Bound

Recall from lecture 1, Hamming proved a bound for binary codes:

Define $\mathrm{Vol}_q(n, r)$ to be volume of ball of radius r in Σ^n, where $|\Sigma| = q$.

Then Hamming claimed $2^k \cdot \mathrm{Vol}_2(n, (d-1)/2) \leq 2^n$.

Asymptotically $R + H_2(\delta/2) \leq 1$.

q-ary generalization:

$q^k \cdot \mathrm{Vol}_q(n, (d-1)/2) \leq q^n$.

Asymptotically $R + H_q(\delta/2) \leq 1$, where $H_q(p) = -p \log_q p - (1-p)\log_q(1-p) + p \log_q(q-1)$.

Question: Are these bounds in the right ballpark?

If bounds are tight, it implies there could be codes of positive rate at $\delta = 1$. Is this feasible? Will rule this out in the next few lectures.

If bounds are in the right ballpark, there exist codes of positive rate and relative distance. Is this feasible? YES! Lets show this.

The random code

Recall the implication of Shannon's theorem: Can correct p fraction of (random) error, with encoding algorithms of rate $1 - H(p)$. Surely this should give a nice code too? Will analyze below.

Code: Pick 2^k random codewords in $\{0,1\}^n$. Lets analyze distance.

The random code

Lets pick $c_1, \ldots, c_K$ at random from $\{0,1\}^n$ and consider the probabilty that they are all pairwise hope they are at distance $d = \delta n$.

Let X_i be the indicator variable for the event that the codeword c_i is at distance less than d from some codeword c_j for $j < i$.

Note that the probability that $X_i = 1$ is at most $(i-1) \cdot 2^{H(\delta)\cdot n}/2^n$.

Thus the probability that there exists an i such that $X_i = 1$ is at most $\sum_{i=1}^{K} (i-1) \cdot 2^{H(\delta)-1\cdot n}$.

The final quantity above is roughly $2^{(2R+H(\delta)-1)\cdot n}$ and thus we have that we can get codes of rate R with relative distance δ provided $2R + H(\delta) < 1$.

A better random code

The bound we have so far only says we can get codes of rate $\frac{1}{2}$ as the relative distance approaches 0. One would hope to do better.

However, we don't know of better ways to estimate either the probability that $X_i = 1$, or the probability that $\{\exists i \mid X_i = 1\}$.

Turns out, a major weakness is in our interpretation of the results. Notice that if $X_i = 1$, it does not mean that the code we found is *totally* bad. It just means that we have to throw out the word c_i from our code. So rather than analyzing the probability that all X_is are 0, we should analyze the probability of the event $\sum_{i=1}^{K} X_i \geq K/2$. If we can bound this probability away from 1 for

some K, then it means that there exist codes with $K/2$ codewords that have distance at least d. Furthermore if the probability that $X_K = 1$ is less than $1/10$, we have that the probability that $\sum_{i=1}^{K} X_i > K/2$ is at most $\frac{1}{5}$ (by Markov's Inequality) and so it suffices to have $E[X_K] = K 2^{(H(\delta)-1)\cdot n} \leq \frac{1}{10}$. Thus, we get that if $R + H(\delta) < 1$ then there exists a code with rate R and distance δ.

In the Problem Set, we will describe many other proofs of this fact.