The Gilbert Varshamov bound

The previous bounds (Singleton, Hamming) we saw for the number of codewords gave an upper bound on the number of codewords. The bounds we are going to discuss give a lower bound. There are the different versions of the Gilbert-Varshamov (G-V) bound.

Proposition. *Let C be an $(n, M, d)_q$ code. If $M \cdot |B(0, d-1)| < q^n$, then there is an $(n, M+1, d)_q$ code. Hence, there is an $(n, M, d)_q$ code for which $M \geq \lceil q^n / |B(0, d-1)| \rceil$ holds.*

Proof. If $M \cdot |B(0, d-1)| < q^n$, then there is a word u, which has distance $\geq d$ from every codeword. Then $C \cup \{u\}$ will have minimum distance $\geq d$.

This observation means that one can construct a code with $M \geq \lceil q^n / |B(0, d-1)| \rceil$ greedily. The next proposition shows an analogous statement for linear codes.

Proposition. *If $|B(0, d-1)| < q^{n-k}$, then there is an $[n, k+1, d]_q$ code.*

Proof. We may assume (by induction on k) that there is an $[n, k, d]_q$ code C. We would like to extend its dimension by one. As before, there is a u which is not in the union of the balls of radius $d-1$ around codewords. Consider $C' =< C, u >$. We have to show that $d(\lambda u + \mu c, c') \geq d$ for every $c, c' \in C$ (and $\lambda \neq 0$). If $d(\lambda u + \mu c, c') = d(u + (\mu/\lambda)c, (1/\lambda)c') \leq d-1$, then we find $c'', c^* \in c$ so that $d(u + c'', c^*) \leq d-1$. But $d(u + c'', c^*) = d(u, c^* - c'') \geq d$, a contradiction. A similar argument shows that $d(\lambda u + \mu c, \lambda' u + \mu' c') \geq d$. So, the $(k+1)$-dimensional linear code $C' =< C, u >$ has minimum distance at least d.

Definition. An (infinite) family $\mathcal{X} = \{C_n\}$ of $[n, k, d]_q$ code is *good* if it has a subsequence $\{C_{n_i}\}$ with $n_i \to \infty$ and $k_i/n_i \geq R > 0$, $d_i/n_i \geq \delta > 0$.

The number δ is called the *relative distance* of the code. So, the previous definition want a subsequence whose information rate is above a positive constant R, relative distance is above a positive constant δ. In order to show the existence of good codes in a more limited family of linear codes, we extend the previous assertion (with a different proof) to certain special families of linear codes.

Theorem (general Gilbert-Varshamov bound)**.** *Assume that we have a family $\mathcal{X}$ of linear $[n, k]_q$ codes which satisfies the following property: $(*)$ for every $0 \neq v$ the number of codes $C \in \mathcal{X}$ containing v is independent from v. In other*

words, there is a constant $c = c(\mathcal{X})$ s.t. for every $0 \neq v$, $|\{C \in \mathcal{X} : v \in C\}| = c$. If $|B(0, d-1)| - 1 < q^{n-k}$, then there is a code $C \in \mathcal{X}$ whose minimum distance is $\geq d$.

Let us first count in two ways the pairs $\{(v, C) : 0 \neq v \in C \in \mathcal{X}\}$. (We impose no restriction on the weight of v.) From left to right it is $(q^n - 1)c$, from right to left it is $|\mathcal{X}|(q^k - 1)$. So we get $c = (q^k - 1)|\mathcal{X}|/(q^n - 1) \leq q^{k-n}|\mathcal{X}|$. Let us call a code $C \in \mathcal{X}$ containing a non-zero vector of weight $\leq d - 1$ bad. The number of such vectors is $|B(0, d-1)| - 1$, hence the number of bad codes is at most

$$(|B(0, d-1)| - 1) \cdot c \leq |\mathcal{X}| q^{k-n}(|B(0, d-1)| - 1).$$

By our assumption the right hand side is strictly less than $|\mathcal{X}|$.

The proof actually shows that whenever $q^{k-n}(|B(0, d-1)| - 1) \leq \varepsilon$, then a random code chosen from $\mathcal{X}$ will have minimum distance $\geq d$ with probability at least $1 - \varepsilon$.

Is there a family which satisfies the extra condition? For example, the set of all linear codes of length n, dimension k do. In this case $c = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q$.

Unfortunately, as we remarked earlier, decoding is difficult, so random codes cannot be decoded efficiently.

Using tricks from analysis (for example, Stirling formula for estimating $n!$), one can determine $|B(0, d-1)| - 1$ if n is large. We will not prove the next theorem.

Theorem. *If $\delta \leq \frac{q-1}{q}$, then*

$$\lim_{n \to \infty} \frac{1}{n} \log_q(|B(0, [n\delta] - 1)|) = -\delta \log_q \delta - (1-\delta) \log_q (1-\delta) - \delta \log_q (q-1) = H_q(\delta).$$

This means that roughly $(|B(0, d-1)| - 1)$ is $q^{nH_q(\delta)}$, if n is large enough. This allows us to prove the asymptotic version of the bounds. If the information rate is R and $R + H_q(\delta) < 1$, then we can achieve that $q^{k-n}(|B(0, d-1)| - 1) \leq \varepsilon$ (for an arbitrary small ε, n has to be large enough). This means that with high probablity (at least $1 - \varepsilon$) a random code from the family will have information rate $\geq R$, and relative distance $\geq \delta$. This is the asymptotic version of the G-V bound.

Theorem (Asymptotic G-V bound)**.** *Assume that $\delta \leq \frac{q-1}{q}$ and $R < 1 - H_g(\delta)$. Then there is a linear code with information rate $\geq R$, and relative distance $\geq \delta$.*

Note that the Singleton bound also has an asymtotic version, which is very simple, it says $\delta \leq 1 - R + 1/n$ or $R \leq 1 - \delta + 1/n$. The asymptotic version of the Hamming bound is also not difficult, it says $R \leq 1 - H_q(\frac{\delta}{2} + o(1)$.