TUTORIAL X

1 q-ary Entropy and Volume of Hamming Balls

q-ary entropy function: Let q be an integer and x be a real number such that $q \geq 2$ and $0 \leq x \leq 1$. Then the q-ary entropy function is defined as follows:

$$H_q(x) = x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x).$$

Volume of a Hamming ball: Let $q \geq 2$ and $n \geq r \geq 1$ be integers. The volume of a Hamming ball of radius r is given by

$$\mathrm{Vol}_q(r,n) = |B_q(\mathbf{0}, r)| = \sum_{i=0}^{r} \binom{n}{i} (q-1)^i.$$

For $0 \leq p \leq 1 - \frac{1}{q}$ real, show that the following bounds hold for large enough n.

1. $\mathrm{Vol}_q(pn, n) \leq q^{nH_q(p)}$.

   A:

$$\begin{aligned}
1 &= (p + (1-p))^n \\
&= \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{pn} \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=pn+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\
&\geq \sum_{i=1}^{pn} \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{pn} \binom{n}{i} (q-1)^i \left(\frac{p}{q-1}\right)^i (1-p)^{n-i} \\
&= \sum_{i=1}^{pn} \binom{n}{i} (q-1)^i (1-p)^n \left(\frac{p}{(q-1)(1-p)}\right)^i \\
&\geq \sum_{i=1}^{pn} \binom{n}{i} (q-1)^i (1-p)^n \left(\frac{p}{(q-1)(1-p)}\right)^{pn} \\
&= \left(\frac{p}{q-1}\right)^{pn} (1-p)^{(1-p)n} \sum_{i=1}^{pn} \binom{n}{i} (q-1)^i \\
&\geq \mathrm{Vol}_q(pn, n) q^{-nH_q(p)}
\end{aligned}$$

2. $\mathrm{Vol}_q(pn, n) \geq q^{nH_q(p) - o(n)}$. (Hint: Use Stirling's approximation)

   A: *Stirling's approximation gives the following bounds on $n!$.*

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\lambda_1(n)} \leq n! \leq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\lambda_2(n)}$$

(We could take, for instance, $\lambda_1(n) = 0$ and $\lambda_2(n) = \frac{1}{12n}$). We have:

$$\begin{aligned}\binom{n}{pn} &= \frac{n!}{(pn)!((1-p)n)!} \\ &> \frac{(n/e)^n}{(pn/e)^{pn}((1-p)n/e)^{(1-p)n}} \cdot \underbrace{\frac{e^{\lambda_1(n)-\lambda_2(pn)-\lambda_2((1-p)n)}}{\sqrt{2\pi p(1-p)n}}}_{\ell(n)} \\ &= \frac{\ell(n)}{p^{pn}(1-p)^{(1-p)n}}\end{aligned}$$

We now derive the wanted lower bound:

$$\begin{aligned}\mathrm{Vol}_q(pn, n) &\geq \binom{n}{pn}(q-1)^{pn} \\ &> \frac{(q-1)^{pn}}{p^{pn}(1-p)^{(1-p)n}} \cdot l(n) \\ &\geq q^{nH_q(p)+\log_q l(n)}\end{aligned}$$

2 Codes Achieving the Gilbert-Varshamov Bound

The purpose of this exercise is to use the probabilistic method to show that a random linear code lies on the Gilbert-Varshamov bound, with high probability.

1. Given a non-zero vector $\mathbf{m} \in \mathbb{F}_q^k$ and a uniformly random $k \times n$ matrix $\mathbf{G}$ over $\mathbb{F}_q$, show that the vector $\mathbf{mG}$ is uniformly distributed over $\mathbb{F}_q^n$.

   A: *Let's denote $\mathbf{m}$ by its coordinates $\mathbf{m} = (m_1, \ldots, m_k)$. The vector $\mathbf{x} = \mathbf{mG}$ is defined by the following equalities:*

   $$\mathbf{x}_i = \sum_{j=1}^{k} m_j g_{j,i}$$

   Thus, for $i \neq j$, b_i and b_j are independent since they depends on disjoint subsets of values of $\mathbf{G}$, which are taken at random. It remains to show that for a given i, the value $\mathbf{x}_i$ is uniformly distributed over $\mathbb{F}_q$. Wlog, we can assume $m_1 \neq 0$ since $\mathbf{m}$ is non-zero, thus we can rewrite $\mathbf{x}_i$ as:

   $$\mathbf{x}_i = m_1 g_{1,i} + \sum_{j=2}^{k} m_j g_{j,i}$$

   Now, for given values of $g_{2,i}, \ldots, g_{k,i}$, the values for $\mathbf{x}_i$ for $g_{1,i} \in \mathbb{F}_q$ are all different. Hence, all values of $\mathbb{F}_q$ for $\mathbf{x}_i$ are equiprobable, thus $\mathbf{x}$ is uniformly distributed of $\mathbb{F}_q^n$.

2. For $k = (1 - H_q(\delta) - \varepsilon)n$, show that there exists a $k \times n$ matrix $\mathbf{G}$ such that

   $$\text{for every } \mathbf{m} \in \mathbb{F}_q^k \setminus \{\mathbf{0}\}, wt(\mathbf{mG}) \geq d$$

   A: *Using question above, for a given $\mathbf{m}$, the following equality holds:*

   $$\mathbf{P}(wt(\mathbf{mG}) < d) = \frac{\mathrm{Vol}_q(d-1, n)}{q^n}$$

Using inequality from previous exercise, this probability is upper bounded by $q^{n(H_q(\delta)-1)}$. By union bound, we have:

$$\begin{aligned}\mathbf{P}(\exists \mathbf{m} \in \mathbb{F}_q^k,\ wt(\mathbf{mG}) < d) &\leq q^k q^{n(H_q(\delta)-1)} \\ &= q^{n(1-H_q(\delta)-\varepsilon)+n(H_q(\delta)-1)} \\ &= q^{-\varepsilon n}\end{aligned}$$

We have $q^{-\varepsilon n} << 1$, thus such a code C does exist.

3. Show that $\mathbf{G}$ has full rank (i.e., it has dimension at least $k = (1 - H_q(\delta) - \varepsilon)n$)

***A:** For a linear code, one has: $d = \min_{c \in C^*} wt(c)$, thus, this code has a distance at least $d > 0$, hence all different $\mathbf{m} \in \mathbb{F}_q^k$ are mapped to different codewords.*

3 Singleton Bound

For every $[n, k, d]$-linear code, show that $k \leq n - d + 1$.

***A:** Let $c_1, c_2, \ldots, c_M$ be the codewords of an $(nk, d)_q$ code C. The singleton bound is equivalent to proving $M \leq q^{n-d+1}$. For a codeword c_i, we define c_i' the prefix of c_i of length $n - d + 1$. We now claim that for any $i \neq j$, we have $c_i' \neq c_j'$. Indeed, if it was the case, we would have $\Delta(c_i, c_j) \leq n - (n - d + 1) = d - 1$, which is impossible since C has distance d. Hence M is the number of prefixes of codewords in C of length $n - d + 1$, which implies that $M \leq q^{n-d+1}$*

4 Weights of Codewords

Let C be an $[n, k, d]$-linear code over $\mathbb{F}_q$. Prove the following.

1. For $q = 2$, either all the codewords have even weight or exactly half have even weight and the rest have odd weight.

   ***A:** Denote by E the set of codewords of C with even weight, and O the set of codewords with odd weight. If $O = \emptyset$, then $E = C$, ie all codewords have even length. Otherwise, there exists $x_0 \in O$. Define the map $\rho : E \to O$ such that $\rho(x) = x + x_0$. Indeed, if x has a even weight, then $x + x_0$ must have an odd weight. Moreover, ρ is an involution, ie $\rho \circ \rho = Id$. Thus, ρ is a bijection and $|E| = |O|$. Since $|C| = |E| + |O|$, we have $|E| = |O| = \frac{|C|}{2}$.*

2. For any q, either all the codewords begin with 0 or exactly a fraction $1/q$ of the codewords begin with 0. In general, for a given position $1 \leq i \leq n$, either all codewords contain 0 at the i-th position or each $\alpha \in \mathbb{F}_q$ appears at the i-th position of exactly $1/q$ of the codewords in C.

   ***A:** Exactly the same kind of proof.*

3. The following inequality holds for the minimum distance d of C.

$$d \leq \frac{n(q-1)q^{k-1}}{q^k - 1}$$

   A: