Homework 3

1. In this problem we will prove a tight upper and lower bound on the size of $\mathsf{Ball}_2(n, r)$ using induction.

   Recall that $\mathsf{Ball}_2(n, r)$ is the number of binary strings with weight at most r. The size of $\mathsf{Ball}_2(n, r)$ is represented by $\mathsf{Vol}_2(n, r)$. And, we have:

   $$\mathsf{Vol}_2(n, r) = \sum_{i=0}^{r} \binom{n}{i}$$

   Recall that $h_2(x) = -x \log x - (1 - x) \log(1 - x)$, for $x \in [0, 1]$. We will represent

   $$H_2(n, r) = \exp(n h_2(r/n)) = \frac{n^n}{r^r (n - r)^{(n-r)}}$$

   We will prove the following statement. For $1 \leqslant r \leqslant n/2$, we have:

   $$\frac{1}{\mathrm{e}} \sqrt{\frac{n}{r(n - r)}} \leqslant \frac{\mathsf{Vol}_2(n, r)}{H_2(n, r)} \leqslant \frac{3}{4}$$

   (a) (5 points) **Base Case.** Prove the statement for $1 = r \leqslant n/2$.

   (b) **Induction.** In the following steps, we will perform the inductive step.

      i. (10 points) **Outlier Case.** Prove the statement for $1 \leqslant r = n/2$, when n is even. You may use the following bound:

      $$\frac{2^n}{\sqrt{2n}} \leqslant \binom{n}{n/2} \leqslant \frac{2^n}{\sqrt{3n/2 + 1}}$$

      ii. (5 points) **Preparation.** Prove that $\mathsf{Vol}_2(n, r) = \mathsf{Vol}_2(n - 1, r - 1) + \mathsf{Vol}_2(n - 1, r)$.

      iii. (2 points) **Sanity Check.** For $2 \leqslant r \leqslant n/2$, prove: If it is not the case that "n is even and $r = n/2$" then $(r - 1) \leqslant (n - 1)/2$ and $r \leqslant (n - 1)/2$. Note that without verifying this, we cannot apply the next inductive step!

      iv. (28 points) **Main Inductive Step.** Assuming that the statement is true for all $1 \leqslant r \leqslant n/2$ such that $n < N$, prove the statement for $n = N$. You may need to prove the following inequalities. For natural numbers a, b, we have

      $$\frac{a^a}{(a - 1)^{(a-1)}} + \frac{b^b}{(b - 1)^{(b-1)}} \leqslant \frac{(a + b)^{(a+b)}}{(a + b - 1)^{(a+b-1)}}$$
      $$\frac{a^{a+1/2}}{(a - 1)^{(a-1/2)}} + \frac{b^{b+1/2}}{(b - 1)^{(b-1/2)}} \geqslant \frac{(a + b)^{(a+b+1/2)}}{(a + b - 1)^{(a+b-1/2)}}$$

2. (5 + 10 points) In this problem we will generalize the Gilbert-Varshamov bound to arbitrary fields.

   For a field $\mathbb{F}$ of size q, let $\mathsf{Ball}_q(n, r)$ be the ball of radius r in $\mathbb{F}^n$ centered at the origin. Let $\mathsf{Vol}_q(n, r)$ be the size of $\mathsf{Ball}_q(n, r)$. Provide an exact expression for $\mathsf{Vol}_q(n, r)$.

   State and prove the Gilbert-Varshamov bound for codes in $\mathbb{F}^n$.

3. (10 + 10 points) In this problem we shall show that a random generator matrix of appropriate dimension (nearly) achieves the Gilbert-Varshamov Bound with high probability. This technique can also be used to show the existence of capacity achieving linear codes for appropriate noisy channels (Shannon's Noisy Channel Coding Theorem).

   Consider the following sampling algorithm.

   Sample (n, k):

   (a) Let $i = 1$

   (b) While $(i \leqslant k)$:

      i. Sample random $v_i \xleftarrow{\$} \{0,1\}^n$

      ii. Increment i

   (c) Let $\mathcal{C}$ be the code spanned by the vectors $\{v_1, \ldots, v_k\}$

   (d) Return the code $\mathcal{C}$

   Given d and t, find as large a value of k as possible such that

   $$\mathbb{P}\left[\text{The code } \mathcal{C} \text{ is an } [n, k, d]_2\text{-code}\right] \geqslant 1 - 2^{-t}$$

   Generalize the algorithm to work for an arbitrary field $\mathbb{F}$ and solve the same problem.

4. (5 + 5 + 5 points) In this problem, we will study some interesting properties of linear codes. These properties are useful in designing secret sharing schemes. In the sequel, we only study binary linear codes. But the result carry over to linear codes over arbitrary fields.

   Let G be the generator matrix of an $[n, k, d]_2$ code. This implies that G is a rank k binary matrix of dimension $k \times n$. The code $\mathcal{C}$ generated by G is the span of all the rows of G. And any non-zero codeword in $\mathcal{C}$ has weight at least d.

   Let $\mathbb{U}_k$ be the uniform distribution over $\{0,1\}^k$. Then, $\mathbb{U}_k \cdot G$ is the uniform distribution over $\mathcal{C}$. We represent this as the joint distribution $(\mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_n)$, where for $1 \leqslant i \leqslant n$, the random variable $\mathbb{C}_i$ is over the sample space $\{0,1\}$.

   (a) Prove that the marginal distribution $\mathbb{C}_i$ is a uniform distribution over $\{0,1\}$ if and only if the i-th column of G, represented by $G_{*,i}$, is non-zero.
   Note that this implies the following. If the column $G_{*,i}$ is the all zero column, then $\mathbb{C}_i$ is 0 with probability 1. Otherwise, $\mathbb{C}_i$ is a uniform random bit.

   (b) For $i, j \in [n]$, let $c_i \in \{0,1\}$ be a bit such that $\mathbb{P}[\mathbb{C}_i = c_i] > 0$. Prove that the conditional distribution $(\mathbb{C}_j | \mathbb{C}_i = c_i)$ is a uniform distribution over $\{0,1\}$ if and only if $G_{*,j}$ is not in the span of $G_{*,i}$.
   Note that this result implies the first result! And, if $G_{*,j} = G_{*,i}$, then the distribution $(\mathbb{C}_j | \mathbb{C}_i = c_i)$ is identical to the distribution that always outputs c_i.

(c) For $i_1, \ldots, i_t \in [n]$, let $(c_{i_1}, \ldots, c_{i_{t-1}})$ be a $(t-1)$-bit string such that

$$\mathbb{P}\left[\mathbb{C}_{i_1} = c_{i_1} \wedge \cdots \wedge \mathbb{C}_{i_{t-1}} = c_{i_{t-1}}\right] > 0$$

Prove that the conditional distribution $(\mathbb{C}_{i_t} | \mathbb{C}_{i_1} = c_{i_1} \wedge \cdots \wedge \mathbb{C}_{i_{t-1}} = c_{i_{t-1}})$ is a uniform distribution over $\{0,1\}$ if and only if G_{*,i_t} is not in the span of $\{G_{*,i_1}, \ldots, G_{*,i_{t-1}}\}$.

Note that this result implies the second result! And, if G_{*,i_t} is in the span of $\{G_{*,i_1}, \ldots, G_{*,i_{t-1}}\}$, then the distribution $(\mathbb{C}_{i_t} | \mathbb{C}_{i_1} = c_{i_1} \wedge \cdots \wedge \mathbb{C}_{i_{t-1}} = c_{i_{t-1}})$ has a deterministic output.