Foreword

This chapter is based on lecture notes from coding theory courses taught by Venkatesan Guruswami at University at Washington and CMU; by Atri Rudra at University at Buffalo, SUNY and by Madhu Sudan at MIT.

This version is dated **August 15, 2014**. For the latest version, please go to

http://www.cse.buffalo.edu/ atri/courses/coding-theory/book/

The material in this chapter is supported in part by the National Science Foundation under CAREER grant CCF-0844796. Any opinions, findings and conclusions or recomendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

Chapter 3

Probability as Fancy Counting and the q-ary Entropy Function

In the first half of this chapter, we will develop techniques that will allow us to answer questions such as

Question 3.0.1. *Does there exist a* $[2,2,1]_2$ *code?*

We note that the answer to the above question is trivially yes: just pick the generator matrix to be the 2×2 identity matrix. However, we will use the above as a simple example to illustrate a powerful technique called the *probabilistic method*.

As the name suggests, the method uses probability. Before we talk more about the probabilistic method, we do a quick review of the basics of probability that we will need in this book.

3.1 A Crash Course on Probability

In this book, we will only consider probability distributions defined over finite spaces. In particular, given a finite domain $\mathbb{D}$, a probability distribution is defined as a function

$$p : \mathbb{D} \to [0,1] \text{ such that } \sum_{x \in \mathbb{D}} p(x) = 1,$$

where $[0,1]$ is shorthand for the interval of all real numbers between 0 and 1. In this book, we will primarily deal with the following special distribution:

Definition 3.1.1 (Uniform Distribution)**.** The *uniform distribution* over $\mathbb{D}$, denoted by $\mathcal{U}_{\mathbb{D}}$, is given by

$$\mathcal{U}_{\mathbb{D}}(x) = \frac{1}{|\mathbb{D}|} \text{ for every } x \in \mathbb{D}.$$

Typically we will drop the subscript when the domain $\mathbb{D}$ is clear from the context.

G	$\mathscr{U}(G)$	V_{00}	V_{01}	V_{10}	V_{11}
$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\frac{1}{16}$	0	0	0	0
$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$	$\frac{1}{16}$	0	1	0	1
$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$	$\frac{1}{16}$	0	1	0	1
$\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$	$\frac{1}{16}$	0	2	0	2
$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$	$\frac{1}{16}$	0	0	1	1
$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$	$\frac{1}{16}$	0	1	1	0
$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	$\frac{1}{16}$	0	1	1	2
$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$	$\frac{1}{16}$	0	2	1	1

G	$\mathscr{U}(G)$	V_{00}	V_{01}	V_{10}	V_{11}
$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$	$\frac{1}{16}$	0	0	1	1
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\frac{1}{16}$	0	1	1	2
$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\frac{1}{16}$	0	1	1	0
$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$	$\frac{1}{16}$	0	2	1	1
$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$	$\frac{1}{16}$	0	0	2	2
$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$	$\frac{1}{16}$	0	1	2	1
$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$	$\frac{1}{16}$	0	1	2	1
$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$\frac{1}{16}$	0	2	2	0

Table 3.1: Uniform distribution over $\mathbb{F}_2^{2\times 2}$ along with values of four random variables.

For example, consider the domain $\mathbb{D} = \mathbb{F}_2^{2\times 2}$, i.e. the set of all 2×2 matrices over $\mathbb{F}_2$. (Note that each such matrix is a generator matrix of some $[2,2]_2$ code.) The first two columns of Table 3.1 list the elements of this $\mathbb{D}$ along with the corresponding probabilities for the uniform distribution.

Typically, we will be interested in a real-valued function defined on $\mathbb{D}$ and how it behaves under a probability distribution defined over $\mathbb{D}$. This is captured by the notion of a random variable:

Definition 3.1.2 (Random Variable)**.** Let $\mathbb{D}$ be a finite domain and $I \subset \mathbb{R}$ be a finite[1] subset. Let p be a probability distribution defined over $\mathbb{D}$. A *random variable* is a function:

$$V : \mathbb{D} \to I.$$

The *expectation* of V is defined as

$$\mathbb{E}[V] = \sum_{x \in \mathbb{D}} p(x) \cdot V(x).$$

For example, given $(i, j) \in \{0,1\}^2$, let V_{ij} denote the random variable $V_{ij}(G) = wt\big((i, j) \cdot G\big)$, for any $G \in \mathbb{F}_2^{2\times 2}$. The last four columns of Table 3.1 list the values of these four random variables.

In this book, we will mainly consider binary random variables, i.e., with $I = \{0,1\}$. In particular, given a predicate or *event* E over $\mathbb{D}$, we will define its *indicator variable* $\mathbb{1}_E$ to be 1 if E is

[1] In general, I need not be finite. However, for this book this definition suffices.

true and 0 if E is false. Sometimes, we will abuse notation and use E instead of $\mathbb{1}_E$. For example, consider the expectations of the four indicator variables:

$$\mathbb{E}\left[\mathbb{1}_{V_{00}=0}\right] = 16 \cdot \frac{1}{16} = 1.$$

$$\mathbb{E}\left[\mathbb{1}_{V_{01}=0}\right] = 4 \cdot \frac{1}{16} = \frac{1}{4}. \tag{3.1}$$

$$\mathbb{E}\left[\mathbb{1}_{V_{10}=0}\right] = 4 \cdot \frac{1}{16} = \frac{1}{4}. \tag{3.2}$$

$$\mathbb{E}\left[\mathbb{1}_{V_{11}=0}\right] = 4 \cdot \frac{1}{16} = \frac{1}{4}. \tag{3.3}$$

3.1.1 Some Useful Results

Before we proceed, we record a simple property of indicator variables that will be useful. (See Exercise 3.1.)

Lemma 3.1.1. *Let E be any event. Then*

$$\mathbb{E}[\mathbb{1}_E] = \Pr[E \textit{ is true}].$$

Next, we state a simple yet useful property of expectation of a sum of random variables:

Proposition 3.1.2 (Linearity of Expectation)**.** *Given random variables $V_1, \ldots, V_m$ defined over the same domain $\mathbb{D}$ and with the same probability distribution p, we have*

$$\mathbb{E}\left[\sum_{i=1}^{m} V_i\right] = \sum_{i=1}^{m} \mathbb{E}[V_i].$$

Proof. For notational convenience, define $V = V_1 + \cdots + V_m$. Thus, we have

$$\mathbb{E}[V] = \sum_{x \in \mathbb{D}} V(x) \cdot p(x) \tag{3.4}$$

$$= \sum_{x \in \mathbb{D}} \left(\sum_{i=1}^{m} V_i(x)\right) \cdot p(x) \tag{3.5}$$

$$= \sum_{i=1}^{m} \sum_{x \in \mathbb{D}} V_i(x) \cdot p(x) \tag{3.6}$$

$$= \sum_{i=1}^{m} \mathbb{E}[V_i]. \tag{3.7}$$

In the equalities above, (3.4) and (3.7) follow from the definition of expectation of a random variable. (3.5) follows from the definition of V and (3.6) follows by switching the order of the two summations. □

As an example, we have

$$\mathbb{E}\left[\mathbb{1}_{V_{01}=0} + \mathbb{1}_{V_{10}=0} + \mathbb{1}_{V_{11}=0}\right] = \frac{3}{4} \tag{3.8}$$

Frequently, we will need to deal with the probability of the "union" of events. We will use the following result to upper bound such probabilities:

Proposition 3.1.3 (Union Bound)**.** *Given m binary random variables $A_1, \ldots, A_m$, we have*

$$\Pr\left[\left(\bigvee_{i=1}^{m} A_i\right) = 1\right] \leq \sum_{i=1}^{m} \Pr\left[A_i = 1\right].$$

Proof. For every $i \in [m]$, define

$$S_i = \{x \in \mathbb{D} | A_i(x) = 1\}.$$

Then we have

$$\Pr\left[\left(\bigvee_{i=1}^{m} A_i\right) = 1\right] = \sum_{x \in \cup_{i=1}^{m} S_i} p(x) \tag{3.9}$$

$$\leq \sum_{i=1}^{m} \sum_{x \in S_i} p(x) \tag{3.10}$$

$$= \sum_{i=1}^{m} \Pr[A_i = 1]. \tag{3.11}$$

In the above, (3.9) and (3.11) follow from the definition of S_i. (3.10) follows from the fact that some of the $x \in \cup_i S_i$ get counted more than once. □

We remark that the union bound is tight when the events are *disjoint*. (In other words, using the notation in the proof above, when $S_i \cap S_j = \emptyset$ for every $i \neq j$.)

As an example, let $A_1 = \mathbb{1}_{V_{01}=0}$, $A_2 = \mathbb{1}_{V_{10}=0}$ and $A_3 = \mathbb{1}_{V_{11}=0}$. Note that in this case the event $A_1 \vee A_2 \vee A_3$ is the same as the event that there exists a non-zero $\mathbf{m} \in \{0,1\}^2$ such that $wt(\mathbf{m} \cdot G) = 0$. Thus, the union bound implies (that under the uniform distribution over $\mathbb{F}_2^{2\times 2}$)

$$\Pr\left[\text{There exists an } \mathbf{m} \in \{0,1\}^2 \setminus \{(0,0)\}, \text{ such that } wt(\mathbf{m}G) = 0\right] \leq \frac{3}{4}. \tag{3.12}$$

Finally, we present two bounds on the probability of a random variable deviating significantly from its expectation. The first bound holds for any random variable:

Lemma 3.1.4 (Markov Bound)**.** *Let V be a non-zero random variable. Then for any $t > 0$,*

$$\Pr[V \geq t] \leq \frac{\mathbb{E}[V]}{t}.$$

In particular, for any $a \geq 1$,

$$\Pr[V \geq a \cdot \mathbb{E}[V]] \leq \frac{1}{a}.$$

Proof. The second bound follows from the first bound by substituting $t = a \cdot \mathbb{E}[V]$. Thus, to complete the proof, we argue the first bound. Consider the following sequence of relations:

$$\mathbb{E}[V] = \sum_{i \in [0,t)} i \cdot \Pr[V = i] + \sum_{i \in [t,\infty)} i \cdot \Pr[V = i] \tag{3.13}$$

$$\geq \sum_{i \geq t} i \cdot \Pr[V = i] \tag{3.14}$$

$$\geq t \cdot \sum_{i \geq t} \Pr[V = i] \tag{3.15}$$

$$= t \cdot \Pr[V \geq t]. \tag{3.16}$$

In the above relations, (3.13) follows from the definition of expectation of a random variable and the fact that V is positive. (3.14) follows as we have dropped some non-negative terms. (3.15) follows by noting that in the summands $i \geq t$. (3.16) follows from the definition of $\Pr[V \geq t]$.

The proof is complete by noting that (3.16) implies the claimed bound. □

The second bound works only for sums of *independent* random variables. We begin by defining independent random variables:

Definition 3.1.3 (Independence)**.** Two random variables A and B are called *independent* if for every a and b in the ranges of A and B, we have

$$\Pr[A = a \wedge B = b] = \Pr[A = a] \cdot \Pr[B = b].$$

For example, for the uniform distribution in Table 3.1, let A denote the bit $G_{0,0}$ and B denote the bit $G_{0,1}$. It can be verified that these two random variables are independent. In fact, it can be verified all the random variables corresponding to the four bits in G are independent random variables. (We'll come to a related comment shortly.)

Another related concept that we will use is that of probability of an event happening conditioned on another event happening:

Definition 3.1.4 (Conditional Probability)**.** Given two events A and B defined over the same domain and probability distribution, we define the probability of A *conditioned on* B as

$$\Pr[A|B] = \frac{\Pr[A \text{ and } B]}{\Pr[B]}.$$

For example, note that

$$\Pr[\mathbb{1}_{V_{01}=1} | G_{0,0} = 0] = \frac{4/16}{1/2} = \frac{1}{2}.$$

The above definition implies that two events A and B are independent if and only if $\Pr[A] = \Pr[A|B]$. We will also use the following result later on in the book (see Exercise 3.2):

Lemma 3.1.5. *For any two events A and B defined on the same domain and the probability distribution:*

$$\Pr[A] = \Pr[A|B] \cdot \Pr[B] + \Pr[A|\neg B] \cdot \Pr[\neg B].$$

Next, we state the deviation bound. (We only state it for sums of binary random variables, which is the form that will be needed in the book.)

Theorem 3.1.6 (Chernoff Bound)**.** *Let $X_1, \ldots, X_m$ be independent binary random variables and define $X = \sum X_i$. Then the* multiplicative *Chernoff bound sates*

$$\Pr[|X - \mathbb{E}(X)| > \varepsilon \mathbb{E}(X)] < e^{-\varepsilon^2 \mathbb{E}(X)/3},$$

and the additive *Chernoff bound states that*

$$\Pr[|X - \mathbb{E}(X)| > \varepsilon m] < e^{-\varepsilon^2 m/2}.$$

We omit the proof, which can be found in any standard textbook on randomized algorithms.

Finally, we present an alternate view of uniform distribution over "product spaces" and then use that view to prove a result that we will use later in the book. Given probability distributions p_1 and p_2 over domains $\mathbb{D}_1$ and $\mathbb{D}_2$ respectively, we define the product distribution $p_1 \times p_2$ over $\mathbb{D}_1 \times \mathbb{D}_2$ as follows: every element $(x, y) \in \mathbb{D}_1 \times \mathbb{D}_2$ under $p_1 \times p_2$ is picked by choosing x from $\mathbb{D}_1$ according to p_1 and y is picked *independently* from $\mathbb{D}_2$ under p_2. This leads to the following observation (see Exercise 3.3).

Lemma 3.1.7. *For any $m \geq 1$, the distribution $\mathcal{U}_{\mathbb{D}_1 \times \mathbb{D}_2 \times \cdots \times \mathbb{D}_m}$ is identical to the distribution $\mathcal{U}_{\mathbb{D}_1} \times \mathcal{U}_{\mathbb{D}_2} \times \cdots \times \mathcal{U}_{\mathbb{D}_m}$.*

For example, the uniform distribution in Table 3.1 can be described equivalently as follows: pick each of the four bits in G independently and uniformly at random from $\{0, 1\}$.

We conclude this section by proving the following result:

Lemma 3.1.8. *Given a non-zero vector $\mathbf{m} \in \mathbb{F}_q^k$ and a uniformly random $k \times n$ matrix G over $\mathbb{F}_q$, the vector $\mathbf{m} \cdot G$ is uniformly distributed over $\mathbb{F}_q^n$.*

Proof. Let the (j, i)th entry in G ($1 \leq j \leq k, 1 \leq i \leq n$) be denoted by g_{ji}. Note that as G is a random $k \times n$ matrix over $\mathbb{F}_q$, by Lemma 3.1.7, each of the g_{ji} is an independent uniformly random element from $\mathbb{F}_q$. Now, note that we would be done if we can show that for every $1 \leq i \leq n$, the ith entry in $\mathbf{m} \cdot G$ (call it b_i) is an independent uniformly random element from $\mathbb{F}_q$. To finish the proof, we prove this latter fact. If we denote $\mathbf{m} = (m_1, \ldots, m_k)$, then $b_i = \sum_{j=1}^{k} m_j g_{ji}$. Note that the disjoint entries of G participate in the sums for b_i and b_j for $i \neq j$. Given our choice of G, this implies that the random variables b_i and b_j are independent. Hence, to complete the proof we need to prove that b_i is a uniformly independent element of $\mathbb{F}_q$. The rest of the proof is a generalization of the argument we used in the proof of Proposition 2.7.1.

Note that to show that b_i is uniformly distributed over $\mathbb{F}_q$, it is sufficient to prove that b_i takes every value in $\mathbb{F}_q$ equally often over all the choices of values that can be assigned to $g_{1i}, g_{2i}, \ldots, g_{ki}$. Now, as $\mathbf{m}$ is non-zero, at least one of the its element is non-zero: without loss of generality assume that $m_1 \neq 0$. Thus, we can write $b_i = m_1 g_{1i} + \sum_{j=2}^{k} m_j g_{ji}$. Now, for every fixed assignment of values to $g_{2i}, g_{3i}, \ldots, g_{ki}$ (note that there are q^{k-1} such assignments), b_i takes a different value for each of the q distinct possible assignments to g_{1i} (this is where we use the assumption that $m_1 \neq 0$). Thus, over all the possible assignments of $g_{1i}, \ldots, g_{ki}$, b_i takes each of the values in $\mathbb{F}_q$ exactly q^{k-1} times, which proves our claim. □

3.2 The Probabilistic Method

The *probabilistic method* is a very powerful method in combinatorics which can be used to show the existence of objects that satisfy certain properties. In this course, we will use the probabilistic method to prove existence of a code $\mathscr{C}$ with certain property $\mathscr{P}$. Towards that end, we define a distribution $\mathscr{D}$ over all possible codes and prove that when $\mathscr{C}$ is chosen according to $\mathscr{D}$:

$$\Pr\left[\mathscr{C} \text{ has property } \mathscr{P}\right] > 0 \text{ or equivalently } \Pr\left[\mathscr{C} \text{ doesn't have property } \mathscr{P}\right] < 1.$$

Note that the above inequality proves the existence of $\mathscr{C}$ with property $\mathscr{P}$.

As an example consider Question 3.0.1. To answer this in the affirmative, we note that the set of all $[2,2]_2$ linear codes is covered by the set of all 2×2 matrices over $\mathbb{F}_2$. Then, we let $\mathscr{D}$ be the uniform distribution over $\mathbb{F}_2^{2\times 2}$. Then by Proposition 2.3.4 and (3.12), we get that

$$\Pr_{\mathscr{U}_{\mathbb{F}_2^{2\times 2}}}[\text{There is no } [2,2,1]_2 \text{ code}] \leq \frac{3}{4} < 1,$$

which by the probabilistic method answers the Question 3.0.1 in the affirmative.

For the more general case, when we apply the probabilistic method, the typical approach will be to define (sub-)properties $P_1, \ldots, P_m$ such that $\mathscr{P} = P_1 \wedge P_2 \wedge P_3 \ldots \wedge P_m$ and show that for every $1 \leq i \leq m$:

$$\Pr\left[\mathscr{C} \text{ doesn't have property } P_i\right] = \Pr\left[\overline{P_i}\right] < \frac{1}{m}.$$

Finally, by the union bound, the above will prove that[2] $\Pr\left[\mathscr{C} \text{ doesn't have property } \mathscr{P}\right] < 1$, as desired.

As an example, an alternate way to answer Question 3.0.1 in the affirmative is the following. Define $P_1 = \mathbb{1}_{V_{01}\geq 1}$, $P_2 = \mathbb{1}_{V_{10}\geq 1}$ and $P_3 = \mathbb{1}_{V_{11}\geq 1}$. (Note that we want a $[2,2]_2$ code that satisfies $P_1 \wedge P_2 \wedge P_3$.) Then, by (3.1), (3.2) and (3.3), we have for $i \in [3]$,

$$\Pr\left[\mathscr{C} \text{ doesn't have property } P_i\right] = \Pr\left[\overline{P_i}\right] = \frac{1}{4} < \frac{1}{3},$$

as desired.

Finally, we mention a special case of the general probabilistic method that we outlined above. In particular, let $\mathscr{P}$ denote the property that the randomly chosen $\mathscr{C}$ satisfies $f(\mathscr{C}) \leq b$. Then we claim (see Exercise 3.4) that $\mathbb{E}[f(C)] \leq b$ implies that $\Pr[\mathscr{C} \text{ has property } \mathscr{P}] > 0$. Note that this implies that $\mathbb{E}[f(C)] \leq b$ implies that there exists a code $\mathscr{C}$ such that $f(C) \leq b$.

3.3 The q-ary Entropy Function

We begin with the definition of a function that will play a central role in many of our combinatorial results.

[2] Note that $\overline{P} = \overline{P_1} \vee \overline{P_2} \vee \cdots \vee \overline{P_m}$.

Definition 3.3.1 (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:

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

Figure 3.1 presents a pictorial representation of the H_q function for the first few values of q. For the special case of $q = 2$, we will drop the subscript from the entropy function and denote

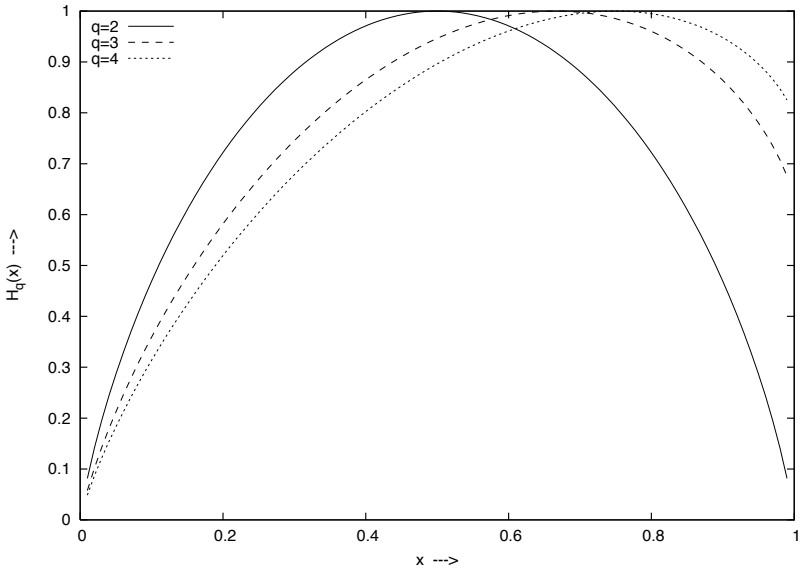

Figure 3.1: A plot of $H_q(x)$ for $q = 2, 3$ and 4. The maximum value of 1 is achieved at $x = 1 - 1/q$.

$H_2(x)$ by just $H(x)$, that is, $H(x) = -x\log x - (1-x)\log(1-x)$, where $\log x$ is defined as $\log_2(x)$ (we are going to follow this convention for the rest of the book).

Under the lens of Shannon's entropy function, $H(x)$ denotes the entropy of the distribution over $\{0,1\}$ that selects 1 with probability x and 0 with probability $1-x$. However, there is no similar analogue for the more general $H_q(x)$. The reason why this quantity will turn out to be so central in this book is that it is very closely related to the "volume" of a Hamming ball. We make this connection precise in the next subsection.

3.3.1 Volume of Hamming Balls

It turns out that in many of our combinatorial results, we will need good upper and lower bounds on the volume of a Hamming ball. Next we formalize the notion of the volume of a Hamming ball:

Definition 3.3.2 (Volume of a Hamming Ball)**.** Let $q \geq 2$ and $n \geq r \geq 1$ be integers. Then the volume of a Hamming ball of radius r is given by

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

The choice of $\mathbf{0}$ as the center for the Hamming ball above was arbitrary: since the volume of a Hamming ball is independent of its center (as is evident from the last equality above), we could have picked any center.

We will prove the following result:

Proposition 3.3.1. *Let $q \geq 2$ be an integer and $0 \leq p \leq 1 - \frac{1}{q}$ be a real. Then for large enough n:*

(i) $Vol_q(pn, n) \leq q^{H_q(p)n}$; *and*

(ii) $Vol_q(pn, n) \geq q^{H_q(p)n - o(n)}$.

Proof. We start with the proof of *(i)*. Consider the following sequence of relations:

$$\begin{aligned}
1 &= (p + (1-p))^n \\
&= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} && (3.17) \\
&= \sum_{i=0}^{pn} \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=pn+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\
&\geq \sum_{i=0}^{pn} \binom{n}{i} p^i (1-p)^{n-i} && (3.18) \\
&= \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i \left(\frac{p}{q-1}\right)^i (1-p)^{n-i} \\
&= \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i (1-p)^n \left(\frac{p}{(q-1)(1-p)}\right)^i \\
&\geq \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i (1-p)^n \left(\frac{p}{(q-1)(1-p)}\right)^{pn} && (3.19) \\
&= \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i \left(\frac{p}{q-1}\right)^{pn} (1-p)^{(1-p)n} && (3.20) \\
&\geq Vol_q(pn, n) q^{-H_q(p)n}. && (3.21)
\end{aligned}$$

In the above, (3.17) follows from the binomial expansion. (3.18) follows by dropping the second sum and (3.19) follows from that facts that $\frac{p}{(q-1)(1-p)} \leq 1$ (as[3] $p \leq 1 - 1/q$) and $pn \geq 1$ (for large enough n). Rest of the steps except (3.21) follow from rearranging the terms. (3.21) follows as $q^{-H_q(p)n} = \left(\frac{p}{q-1}\right)^{pn} (1-p)^{(1-p)n}$.

(3.21) implies that

$$1 \geq Vol_q(pn, n) q^{-H_q(p)n},$$

which proves *(i)*.

[3] Indeed, note that $\frac{p}{(q-1)(1-p)} \leq 1$ is true if $\frac{p}{1-p} \leq \frac{q-1}{1}$, which in turn is true if $p \leq \frac{q-1}{q}$, where the last step follows from Lemma B.2.1.

We now turn to the proof of part *(ii)*. For this part, we will need Stirling's approximation for $n!$ (Lemma B.1.2).

By the Stirling's approximation, we have the following inequality:

$$\begin{aligned}\binom{n}{pn} &= \frac{n!}{(pn)!((1-p)n)!} \\ &> \frac{(n/e)^n}{(pn/e)^{pn}((1-p)n/e)^{(1-p)n}} \cdot \frac{1}{\sqrt{2\pi p(1-p)n}} \cdot e^{\lambda_1(n)-\lambda_2(pn)-\lambda_2((1-p)n)} \\ &= \frac{1}{p^{pn}(1-p)^{(1-p)n}} \cdot \ell(n), \end{aligned} \tag{3.22}$$

where $\ell(n) = \frac{e^{\lambda_1(n)-\lambda_2(pn)-\lambda_2((1-p)n)}}{\sqrt{2\pi p(1-p)n}}$.

Now consider the following sequence of relations that complete the proof:

$$Vol_q(pn,n) \geq \binom{n}{pn}(q-1)^{pn} \tag{3.23}$$

$$> \frac{(q-1)^{pn}}{p^{pn}(1-p)^{(1-p)n}} \cdot \ell(n) \tag{3.24}$$

$$\geq q^{H_q(p)n-o(n)}. \tag{3.25}$$

In the above (3.23) follows by only looking at one term. (3.24) follows from (3.22) while (3.25) follows from the definition of $H_q(\cdot)$ and the fact that for large enough n, $\ell(n)$ is $q^{-o(n)}$. □

Next, we consider how the q-ary entropy function behaves for various ranges of its parameters.

3.3.2 Other Properties of the q-ary Entropy function

We begin by recording the behavior of q-ary entropy function for large q.

Proposition 3.3.2. *For small enough ε, $1 - H_q(\rho) \geq 1 - \rho - \varepsilon$ for every $0 < \rho \leq 1 - 1/q$ if and only if q is $2^{\Omega(1/\varepsilon)}$.*

Proof. We first note that by definition of $H_q\rho)$ and $H(\rho)$,

$$\begin{aligned} H_q(\rho) &= \rho \log_q(q-1) - \rho \log_q \rho - (1-\rho)\log_q(1-\rho) \\ &= \rho \log_q(q-1) + H(\rho)/\log_2 q. \end{aligned}$$

Now if $q \geq 2^{1/\varepsilon}$, we get that

$$H_q(\rho) \leq \rho + \varepsilon.$$

as $\log_q(q-1) \leq 1$ and $H(\rho) \leq 1$. Thus, we have argued that for $q \geq 2^{1/\varepsilon}$, we have $1 - H_q(\rho) \geq 1 - \rho - \varepsilon$, as desired.

Next, we consider the case when $q = 2^{o(1/\varepsilon)}$. We begin by claiming that for small enough ε,

$$\text{if } q \geq 1/\varepsilon^2 \text{ then } \log_q(q-1) \geq 1-\varepsilon.$$

Indeed, $\log_q(q-1) = 1 + (1/\ln q)\ln(1-1/q) = 1 - O\left(\frac{1}{q \ln q}\right)$,[4] which is at least $1-\varepsilon$ for $q \geq 1/\varepsilon^2$ (and small enough ε).

Finally, if $q = 2^{o\left(\frac{1}{\varepsilon}\right)}$, then for fixed ρ,

$$H(\rho)/\log q = \varepsilon \cdot \omega(1).$$

Then for $q = 2^{o\left(\frac{1}{\varepsilon}\right)}$ (but $q \geq 1/\varepsilon^2$) we have

$$\rho \log_q(q-1) + H(\rho)/\log q \geq \rho - \varepsilon + \varepsilon \cdot \omega(1) > \rho + \varepsilon,$$

which implies that

$$1 - H_q(\rho) < 1 - \rho - \varepsilon,$$

as desired. For $q \leq 1/\varepsilon^2$, Lemma 3.3.3 shows that $1 - H_q(\rho) \leq 1 - H_{1/\varepsilon^2}(\rho) < 1 - \rho - \varepsilon$, as desired. □

We will also be interested in how $H_q(x)$ behaves for fixed x and increasing q:

Lemma 3.3.3. *Let $q \geq 2$ be an integer and let $0 \leq \rho \leq 1 - 1/q$, then for any real $m \geq 1$ such that*

$$q^{m-1} \geq \left(1 + \frac{1}{q-1}\right)^{q-1}, \tag{3.26}$$

we have

$$H_q(\rho) \geq H_{q^m}(\rho).$$

Proof. Note that $H_q(0) = H_{q^m}(0) = 0$. Thus, for the rest of the proof we will assume that $\rho \in (0, 1-1/q]$.

As observed in the proof of Proposition 3.3.2, we have

$$H_q(\rho) = \rho \cdot \frac{\log(q-1)}{\log q} + H(\rho) \cdot \frac{1}{\log q}.$$

Using this, we obtain

$$H_q(\rho) - H_{q^m}(\rho) = \rho\left(\frac{\log(q-1)}{\log q} - \frac{\log(q^m-1)}{m \log q}\right) + H(\rho)\left(\frac{1}{\log q} - \frac{1}{m \log q}\right).$$

The above in turn implies that

$$\frac{1}{\rho} \cdot m \log q \cdot (H_q(\rho) - H_{q^m}(\rho)) = \log(q-1)^m - \log(q^m-1) + \frac{H(\rho)}{\rho}(m-1)$$

[4] The last equality follows from the fact that by Lemma B.2.2, for $0 < x < 1$, $\ln(1-x) = -O(x)$.

$$\geq \log(q-1)^m - \log(q^m-1) + \frac{H(1-1/q)}{1-1/q}(m-1) \tag{3.27}$$

$$= \log(q-1)^m - \log(q^m-1) + (m-1)\left(\log\frac{q}{q-1} + \frac{\log q}{q-1}\right)$$

$$= \log\left(\frac{(q-1)^m}{q^m-1}\cdot\left(\frac{q}{q-1}\right)^{m-1}\cdot q^{\frac{m-1}{q-1}}\right)$$

$$= \log\left(\frac{(q-1)\cdot q^{m-1}\cdot q^{\frac{m-1}{q-1}}}{q^m-1}\right)$$

$$\geq 0 \tag{3.28}$$

In the above (3.27) follows from the fact that $H(\rho)/\rho$ is decreasing[5] in ρ and that $\rho \leq 1-1/q$. (3.28) follows from the the claim that

$$(q-1)\cdot q^{\frac{m-1}{q-1}} \geq q.$$

Indeed the above follows from (3.26).

Finally, note that (3.28) completes the proof. □

Since $(1+1/x)^x \leq e$ (by Lemma B.2.3), we also have that (3.26) is also satisfied for $m \geq 1 + \frac{1}{\ln q}$. Further, we note that (3.26) is satisfied for every $m \geq 2$ (for any $q \geq 3$), which leads to the following (also see Exercise 3.5):

Corollary 3.3.4. *Let $q \geq 3$ be an integer and let $0 \leq \rho \leq 1-1/q$, then for any $m \geq 2$, we have*

$$H_q(\rho) \geq H_{q^m}(\rho).$$

Next, we look at the entropy function when its input is very close to 1.

Proposition 3.3.5. *For small enough $\varepsilon > 0$,*

$$H_q\left(1-\frac{1}{q}-\varepsilon\right) \leq 1 - c_q\varepsilon^2,$$

where c_q is a constant that only depends on q.

Proof. The intuition behind the proof is the following. Since the derivative of $H_q(x)$ is zero at $x = 1-1/q$, in the Taylor expansion of $H_q(1-1/q-\varepsilon)$ the ε term will vanish. We will now make this intuition more concrete. We will think of q as fixed and $1/\varepsilon$ as growing. In particular, we will assume that $\varepsilon < 1/q$. Consider the following equalities:

$$H_q(1-1/q-\varepsilon) = -\left(1-\frac{1}{q}-\varepsilon\right)\log_q\left(\frac{1-1/q-\varepsilon}{q-1}\right) - \left(\frac{1}{q}+\varepsilon\right)\log_q\left(\frac{1}{q}+\varepsilon\right)$$

[5] Indeed, $H(\rho)/\rho = \log(1/\rho) - (1/\rho - 1)\log(1-\rho)$. Note that the first term is deceasing in ρ. We claim that the second term is also decreasing in ρ– this e.g. follows from the observation that $-(1/\rho-1)\ln(1-\rho) = (1-\rho)(1+\rho/2!+\rho^2/3!+\cdots) = 1-\rho/2-\rho^2(1/2-1/3!)-\cdots$ is also decreasing in ρ.

$$
\begin{aligned}
&= -\log_q\left(\frac{1}{q}\left(1-\frac{\varepsilon q}{q-1}\right)\right)+\left(\frac{1}{q}+\varepsilon\right)\log_q\left(\frac{1-(\varepsilon q)/(q-1)}{1+\varepsilon q}\right)\\
&= 1-\frac{1}{\ln q}\left[\ln\left(1-\frac{\varepsilon q}{q-1}\right)-\left(\frac{1}{q}+\varepsilon\right)\ln\left(\frac{1-(\varepsilon q)/(q-1)}{1+\varepsilon q}\right)\right]\\
&= 1+o(\varepsilon^2)-\frac{1}{\ln q}\left[-\frac{\varepsilon q}{q-1}-\frac{\varepsilon^2 q^2}{2(q-1)^2}-\left(\frac{1}{q}+\varepsilon\right)\left(-\frac{\varepsilon q}{q-1}\right.\right.\\
&\qquad\left.\left.-\frac{\varepsilon^2 q^2}{2(q-1)^2}-\varepsilon q+\frac{\varepsilon^2 q^2}{2}\right)\right] \qquad (3.29)\\
&= 1+o(\varepsilon^2)-\frac{1}{\ln q}\left[-\frac{\varepsilon q}{q-1}-\frac{\varepsilon^2 q^2}{2(q-1)^2}\right.\\
&\qquad\left.-\left(\frac{1}{q}+\varepsilon\right)\left(-\frac{\varepsilon q^2}{q-1}+\frac{\varepsilon^2 q^3(q-2)}{2(q-1)^2}\right)\right]\\
&= 1+o(\varepsilon^2)-\frac{1}{\ln q}\left[-\frac{\varepsilon^2 q^2}{2(q-1)^2}+\frac{\varepsilon^2 q^2}{q-1}-\frac{\varepsilon^2 q^2(q-2)}{2(q-1)^2}\right] \qquad (3.30)\\
&= 1-\frac{\varepsilon^2 q^2}{2\ln q(q-1)}+o(\varepsilon^2)\\
&\le 1-\frac{\varepsilon^2 q^2}{4\ln q(q-1)}
\end{aligned}
$$

(3.31)

(3.29) follows from the fact that for $|x|<1$, $\ln(1+x)=x-x^2/2+x^3/3-\ldots$ (Lemma B.2.2) and by collecting the ε^3 and smaller terms in $o(\varepsilon^2)$. (3.30) follows by rearranging the terms and by absorbing the ε^3 terms in $o(\varepsilon^2)$. The last step is true assuming ε is small enough. □

Next, we look at the entropy function when its input is very close to 0.

Proposition 3.3.6. *For small enough $\varepsilon>0$,*

$$H_q(\varepsilon)=\Theta\left(\frac{1}{\log q}\cdot\varepsilon\log\left(\frac{1}{\varepsilon}\right)\right).$$

Proof. By definition

$$H_q(\varepsilon)=\varepsilon\log_q(q-1)+\varepsilon\log_q(1/\varepsilon)+(1-\varepsilon)\log_q(1/(1-\varepsilon)).$$

Since all the terms in the RHS are positive we have

$$H_q(\varepsilon)\ge\varepsilon\log(1/\varepsilon)/\log q. \qquad (3.32)$$

Further, by Lemma B.2.2, $(1-\varepsilon)\log_q(1/(1-\varepsilon))\le 2\varepsilon/\ln q$ for small enough ε. Thus, this implies that

$$H_q(\varepsilon)\le\frac{2+\ln(q-1)}{\ln q}\cdot\varepsilon+\frac{1}{\ln q}\cdot\varepsilon\ln\left(\frac{1}{\varepsilon}\right). \qquad (3.33)$$

(3.32) and (3.33) proves the claimed bound.

□

We will also work with the inverse of the q-ary entropy function. Note that $H_q(\cdot)$ on the domain $[0, 1-1/q]$ is an bijective map into $[0,1]$. Thus, we define $H_q^{-1}(y) = x$ such that $H_q(x) = y$ and $0 \le x \le 1-1/q$. Finally, we will need the following lower bound.

Lemma 3.3.7. *For every $0 \le y \le 1-1/q$ and for every small enough $\varepsilon > 0$,*

$$H_q^{-1}(y - \varepsilon^2/c_q') \ge H_q^{-1}(y) - \varepsilon,$$

where $c_q' \ge 1$ is a constant that depends only on q.

Proof. It is easy to check that $H_q^{-1}(y)$ is a strictly increasing convex function in the range $y \in [0,1]$. This implies that the derivative of $H_q^{-1}(y)$ increases with y. In particular, $(H_q^{-1})'(1) \ge (H_q^{-1})'(y)$ for every $0 \le y \le 1$. In other words, for every $0 < y \le 1$, and (small enough) $\delta > 0$, $\frac{H_q^{-1}(y) - H_q^{-1}(y-\delta)}{\delta} \le \frac{H_q^{-1}(1) - H_q^{-1}(1-\delta)}{\delta}$. Proposition 3.3.5 along with the facts that $H_q^{-1}(1) = 1-1/q$ and H_q^{-1} is increasing completes the proof if one picks $c_q' = \max(1, 1/c_q)$ and $\delta = \varepsilon^2/c_q'$. □

3.4 Exercises

Exercise 3.1. Prove Lemma 3.1.1.

Exercise 3.2. Prove Lemma 3.1.5.

Exercise 3.3. Prove Lemma 3.1.7.

Exercise 3.4. Let $\mathscr{P}$ denote the property that the randomly chosen $\mathscr{C}$ satisfies $f(\mathscr{C}) \le b$. Then $\mathbb{E}[f(C)] \le b$ implies that $\Pr[\mathscr{C} \text{ has property } \mathscr{P}] > 0$.

Exercise 3.5. Show that for any $Q \ge q \ge 2$ and $\rho \le 1-1/q$, we have $H_Q(\rho) \le H_q(\rho)$.

3.5 Bibliographic Notes

Shannon was one of the very early adopters of probabilistic method (and we will see one such use in Chapter 6). Later, the probabilistic method was popularized Erdős. For more on probabilistic method, see the book by Alon and Spencer [1].

Proofs of various concentration bounds can e.g. be found in [13].