Error Correcting Codes: Combinatorics, Algorithms and Applications	**(Fall 2007)**

Lecture 9: Converse of Shannon's Capacity Theorem

September 17, 2007

Lecturer: Atri Rudra *Scribe: Thanh-Nhan Nguyen & Atri Rudra*

In the last lecture, we stated Shannon's capacity theorem for the BSC, which we restate here:

Theorem 0.1. *Let $0 \le p < 1/2$ be a real number. For every $0 < \varepsilon \le 1/2 - p$, the following statements are true for large enough integer n:*

(i) There exists a real $\delta > 0$, an encoding function $E : \{0,1\}^k \to \{0,1\}^n$, and a decoding function $D : \{0,1\}^n \to \{0,1\}^k$, where $k \le \lfloor (1 - H(p+\varepsilon))n \rfloor$ such that the following holds for every $\mathbf{m} \in \{0,1\}^k$:

$$\underset{\textit{noise } \mathbf{e} \textit{ of } BSC_p}{Pr}[D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \le 2^{-\delta n}.$$

(ii) If $k \ge \lceil (1 - H(p) + \varepsilon)n \rceil$ then for every encoding and decoding functions $E : \{0,1\}^k \to \{0,1\}^n$ and $D : \{0,1\}^n \to \{0,1\}^k$ the following is true for some $\mathbf{m} \in \{0,1\}^k$:

$$\underset{\textit{noise } \mathbf{e} \textit{ of } BSC_p}{Pr}[D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \ge 1/2.$$

In today's lecture, we will prove part *(ii)* of Theorem 0.1.

1 Preliminaries

Before we begin with the proof we will need a few results, which we discuss first.

1.1 Chernoff Bound

Chernoff bound states a bound on the tail of a certain distribution that will be useful for us. Here we state the version of the Chernoff bound that we will need.

Proposition 1.1. *For $i = 1, \cdots, n$, let X_i be a binary random variable that takes a value of 1 with probability p and a value of 0 with probability $1-p$. Then the following bounds are true:*

(i) $Pr\left[\sum_{i=1}^n X_i \ge (1+\varepsilon)pn\right] \le e^{-\varepsilon^2 pn/3}$

(ii) $Pr\left[\sum_{i=1}^n X_i \le (1-\varepsilon)pn\right] \le e^{-\varepsilon^2 pn/3}$

Note that the expectation of the sum $\sum_{i=1}^n X_i$ is pn. The bound above states that the probability mass is tightly concentrated around the mean.

1.2 Volume of Hamming Balls

We will also need good upper and lower bounds on the volume of a Hamming ball. Recall that $Vol_q(\mathbf{0}, pn) = |B_q(\mathbf{0}, \rho n)| = \sum_{i=0}^{pn} \binom{n}{i}(q-1)^i$. We will prove the following result:

Proposition 1.2. *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) $Vol_q(\mathbf{0}, pn) \leq q^{H_q(p)n}$

(ii) $Vol_q(\mathbf{0}, pn) \geq q^{H_q(p)n - o(n)}$

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

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

$$
\begin{aligned}
1 &= (p + (1-p))^n \\
&= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} && (1)\\
&\geq \sum_{i=0}^{pn} \binom{n}{i} p^i (1-p)^{n-i} && (2)\\
&= \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i \left(\frac{p}{q-1}\right)^i (1-p)^{n-i} \\
&= \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i (1-p)^n \left(\frac{p}{(q-1)(1-p)}\right)^i \\
&\geq \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i (1-p)^n \left(\frac{p}{(q-1)(1-p)}\right)^{pn} && (3)\\
&= \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i \left(\frac{p}{q-1}\right)^{pn} (1-p)^{(1-p)n}. && (4)
\end{aligned}
$$

In the above, (1) follows from the binomial expansion. (2) follows by dropping some terms from the summation and (3) follows from that facts that $\frac{p}{(q-1)(1-p)} \leq 1$ (as $q \geq 2$ and $p \leq 1/2$) and $pn \geq 1$ (for large enough n). Rest of the steps follow from rearranging the terms.

As $q^{-H_q(p)n} = \left(\frac{p}{q-1}\right)^{pn} (1-p)^{(1-p)n}$, (4) implies that

$$1 \geq Vol_q(\mathbf{0}, pn) q^{-H_q(p)n},$$

which proves *(i)*.

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

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

where

$$\lambda_1(n) = \frac{1}{12n+1} \text{ and } \lambda_2(n) = \frac{1}{12n}.$$

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

$$\begin{aligned}
\binom{n}{pn} &= \frac{n!}{(pn)!((1-p)n)!} \\
&> \frac{(n/e)^n}{(pn/e)^{pn}((1-p)n/e)^{(1-p)n}} \cdot \frac{1}{\sqrt{2\pi p(1-p)n}} \cdot e^{\lambda_1(n)-\lambda_2(pn)-\lambda_2((1-p)n)} \\
&= \frac{1}{p^{pn}(1-p)^{(1-p)n}} \cdot \ell(n), \qquad (5)
\end{aligned}$$

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:

$$\begin{aligned}
Vol_q(\mathbf{0}, pn) &\geq \binom{n}{pn}(q-1)^{pn} \qquad (6) \\
&> \frac{(q-1)^{pn}}{p^{pn}(1-p)^{(1-p)n}} \cdot \ell(n) \qquad (7) \\
&\geq q^{H_q(p)n-o(n)}. \qquad (8)
\end{aligned}$$

In the above (6) follows by only looking at one term. (7) follows from (5) while (8) follows from the definition of $H_q(\cdot)$ and the fact that for large enough n, $\ell(n)$ is $q^{-o(n)}$. □

2 Converse of Shannon's Capacity Theorem for BSC

We will now prove part *(ii)* of Theorem 0.1: the proof of the other part will be done in the next lecture.

First, we note that there is nothing to prove if $p = 0$, so for the rest of the proof we will assume that $p > 0$. For the sake of contradiction, assume that the following holds for *every* $\mathbf{m} \in \{0,1\}^k$:

$$\Pr_{\text{noise } \mathbf{e} \text{ of } BSC_p}[D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \leq 1/2.$$

Fix an arbitrary message $\mathbf{m} \in \{0,1\}^k$. Define $D_{\mathbf{m}}$ to be the set of received words that are decoded to $\mathbf{m}$ by D, that is,

$$D_{\mathbf{m}} = \{\mathbf{y} | D(\mathbf{y}) = \mathbf{m}\}.$$

Note that by our assumption, the following is true (where from now on we omit the explicit dependence of the probability on the BSC_p noise for clarity):

$$Pr[E(\mathbf{m}) + \mathbf{e} \notin D_{\mathbf{m}}] \leq 1/2. \qquad (9)$$

Further, by the Chernoff bound,

$$Pr[E(\mathbf{m}) + \mathbf{e} \notin S_{\mathbf{m}}] \le 2^{-\Omega(\gamma^2 n)}, \tag{10}$$

where $S_{\mathbf{m}}$ is the shell of radius $[(1-\gamma)pn, (1+\gamma)pn]$ around $E(\mathbf{m})$, that is, $S_{\mathbf{m}} = B_2(E(\mathbf{m}), (1+\gamma)pn) \setminus B_2(E(\mathbf{m}), (1-\gamma)pn)$. (We will set $\gamma > 0$ in terms of ε and p at the end of the proof.)

(9) and (10) along with the union bound imply the following:

$$Pr\left[E(\mathbf{m}) + \mathbf{e} \in D_{\mathbf{m}} \cap S_{\mathbf{m}}\right] \ge \frac{1}{2} - 2^{-\Omega(\gamma^2 n)} \ge \frac{1}{4}, \tag{11}$$

where the last inequality holds for large enough n. Next we upper bound the probability above to obtain a lower bound on $|D_{\mathbf{m}} \cap S_{\mathbf{m}}|$.

It is easy to see that

$$Pr\left[E(\mathbf{m}) + \mathbf{e} \in D_{\mathbf{m}} \cap S_{\mathbf{m}}\right] \le |D_{\mathbf{m}} \cap S_{\mathbf{m}}| \cdot p_{max},$$

where

$$p_{max} = \max_{\mathbf{y} \in S_{\mathbf{m}}} Pr[E(\mathbf{m}) + \mathbf{e} = \mathbf{y}] = \max_{d \in [(1-\gamma)pn, (1+\gamma)pn]} p^d (1-p)^{n-d}.$$

It is easy to check that $p^d(1-p)^{n-d}$ is decreasing in d for $p \le 1/2$. Thus, we have

$$p_{max} = p^{(1-\gamma)pn}(1-p)^{n-(1-\gamma)pn} = \left(\frac{1-p}{p}\right)^{\gamma pn} \cdot p^{pn}(1-p)^{(1-p)n} = \left(\frac{1-p}{p}\right)^{\gamma pn} 2^{-nH(p)}.$$

Thus, we have shown that

$$Pr\left[E(\mathbf{m}) + \mathbf{e} \in D_{\mathbf{m}} \cap S_{\mathbf{m}}\right] \le |D_{\mathbf{m}} \cap S_{\mathbf{m}}| \cdot \left(\frac{1-p}{p}\right)^{\gamma pn} 2^{-nH(p)},$$

which by (11) implies that

$$|D_{\mathbf{m}} \cap S| \ge \frac{1}{4} \cdot \left(\frac{1-p}{p}\right)^{-\gamma pn} 2^{nH(p)}. \tag{12}$$

Next, we consider the following sequence of relations:

$$\begin{aligned}
2^n &= \sum_{\mathbf{m} \in \{0,1\}^k} |D_{\mathbf{m}}| && (13)\\
&\ge \sum_{\mathbf{m} \in \{0,1\}^k} |D_{\mathbf{m}} \cap S| \\
&\ge \frac{1}{4}\left(\frac{1-p}{p}\right)^{-\gamma pn} \sum_{\mathbf{m} \in \{0,1\}^k} 2^{H(p)n} && (14)\\
&= 2^{k-2} 2^{H(p)n - \gamma p \log(1/p-1)n} \\
&> 2^{k+H(p)n-\varepsilon n}. && (15)
\end{aligned}$$

In the above (13) follows from the fact that for $\mathbf{m}_1 \neq \mathbf{m}_2$, $D_{\mathbf{m}_1}$ and $D_{\mathbf{m}_2}$ are disjoint. (14) follows from (12). (15) follows for large enough n and if we pick $\gamma = \frac{\varepsilon}{2p\log\left(\frac{1}{p}-1\right)}$. (Note that as $0 < p < 1/2$, $\gamma = \Theta(\varepsilon)$.)

(15) implies that $k < (1 - H(p) + \varepsilon)n$, which is a contradiction. The proof of part *(ii)* of Theorem 0.1 is complete.

Remark 2.1. *It can be verified that the proof above can also work if the decoding error probability is bounded by* $2^{-\beta n}$ *(instead of the* $1/2$ *in part (ii) of Theorem 0.1) for small enough* $\beta = \beta(\varepsilon) > 0$.