

# The Noisy Coding Theorem

Michael E. O'Sullivan  
mosulliv@math.sdsu.edu  
www-rohan.sdsu.edu/~mosulliv

April 20, 2004

## 1 Preliminaries

**Theorem 1.1 (Chebyshev's inequality).** *Let  $X$  be a discrete random variable with range  $\mathbb{R}$  and let  $\mu = \mathbb{E}(X)$  and  $\sigma = \sqrt{\text{Var}(X)}$ . Then for any  $c \geq 0$ ,*

$$P(|X - \mu| > c) < \frac{\sigma^2}{c^2}$$

PROOF: We may assume  $\mu = 0$  (just set  $Y = X - \mu$ , then  $\text{Var}(X) = \text{Var}(Y)$ ). By definition,

$$\begin{aligned}\sigma^2 &= \text{Var}(X) \\ &= \sum_{|x_i| \leq c} p_i x_i^2 + \sum_{|x_i| > c} p_i x_i^2 \\ &\geq \sum_{|x_i| > c} p_i x_i^2 \\ &> c^2 P(|X| > c)\end{aligned}$$

So  $P(|X| > c) < \sigma^2/c^2$ . □

**Proposition 1.2.** *Let  $S_n = X_1 + \dots + X_n$  be the sum of independent discrete random variables. Then  $\text{Var } S_n = \sum_{i=1}^n \text{Var } X_i$ .*

**Theorem 1.3 (Law of Large Numbers).** *Let  $S_n = X_1 + \dots + X_n$  be the sum of independent Bernoulli random variables with  $P(X_i = 1) = \alpha$  for all  $i$ . For any  $\delta > 0$ ,*

$$P\left(\left|\frac{S_n}{n} - \alpha\right| > \delta\right) < \frac{\alpha(1-\alpha)}{\delta^2 n}$$

PROOF: We have  $\mu(S_n) = \sum_{i=1}^n \mu(X_i) = n\alpha$ . Also  $\text{Var}(S_n) = n \text{Var}(X) = n\alpha(1-\alpha)$ . Applying Chebyshev's inequality,

$$P(|S_n - n\alpha| > \delta n) < \frac{n\alpha(1-\alpha)}{\delta^2 n^2}$$

The left hand side is equal to  $P\left(\left|\frac{S_n}{n} - \alpha\right| > \delta\right)$ , so we get the result. □

**Definition 1.4.** The entropy of a discrete random variable is

$$H(X) = - \sum P(X = x) \log(P(X = x)).$$

The entropy of a discrete random variable  $X$  has nothing to do with the specific real values taken by  $X$ . It only depends on the probabilities

Let  $\alpha \in [0, 1]$ . we define  $h(\alpha)$  to be the entropy of a random variable taking two distinct values, with probabilities  $\alpha$ , and  $1 - \alpha$ .

$$\begin{aligned} h(\alpha) &= -\alpha \log(\alpha) - (1 - \alpha) \log(1 - \alpha) \\ &= \log(\alpha^{-\alpha} (1 - \alpha)^{-(1-\alpha)}) \end{aligned}$$

**Proposition 1.5 (The Tail Inequality).** *Let  $0 < \alpha < 1/2$ . Then*

$$\sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} < 2^{nh(\alpha)}$$

PROOF:

$$\begin{aligned} 1 &= (\alpha + (1 - \alpha))^n \\ &= \sum_{k=0}^n \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} \\ &> \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} \end{aligned}$$

Since  $0 < \alpha < 1/2$ ,  $\alpha^k (1 - \alpha)^{n-k} > \alpha^r (1 - \alpha)^{n-r}$  when  $r > k$ . Therefore

$$\begin{aligned} \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} &> \alpha^{\alpha n} (1 - \alpha)^{n - \alpha n} \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} \\ &= (\alpha^\alpha (1 - \alpha)^{1-\alpha})^n \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} \\ &= 2^{-nh(\alpha)} \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} \end{aligned}$$

Thus we have the result,  $\sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} < 2^{nh(\alpha)}$ . □

**Proposition 1.6.** *Let  $T$  be a finite set and  $S \subseteq T$ . Let  $s = |S|$  and  $t = |T|$ . The proportion of ways to choose  $r$  objects from  $T$  and get at least one element of  $S$  is upper bounded by  $r \frac{s}{t-r}$ .*

PROOF: The desired proportion is

$$\begin{aligned}
\frac{\binom{t}{r} - \binom{t-s}{r}}{\binom{t}{r}} &= 1 - \frac{\frac{(t-s)!}{r!(t-s-r)!}}{\frac{t!}{r!(t-r)!}} \\
&= 1 - \frac{(t-s)(t-1-s)(t-2-s)\cdots(t-(r-1)-s)}{t(t-1)(t-2)\cdots(t-(r-1))} \\
&= 1 - \left(1 - \frac{s}{t}\right)\left(1 - \frac{s}{t-1}\right)\cdots\left(1 - \frac{s}{t-r+1}\right)
\end{aligned}$$

For  $k \leq r-1 \leq t$  we have

$$\begin{aligned}
t-k &\geq t-r+1 \geq 0 \\
\frac{s}{t-k} &\leq \frac{s}{t-r+1} \\
1 - \frac{s}{t-k} &\geq 1 - \frac{s}{t-r+1} \\
\prod_{k=0}^{r-1} \left(1 - \frac{s}{t-k}\right) &\geq \prod_{k=0}^{r-1} \left(1 - \frac{s}{t-r+1}\right) \\
1 - \prod_{k=0}^{r-1} \left(1 - \frac{s}{t-k}\right) &\leq 1 - \left(1 - \frac{s}{t-r+1}\right)^r
\end{aligned}$$

Thus we have

$$\begin{aligned}
1 - \left(1 - \frac{s}{t}\right)\left(1 - \frac{s}{t-1}\right)\cdots\left(1 - \frac{s}{t-r+1}\right) &\leq 1 - \left(1 - \frac{s}{t-r+1}\right)^r \\
&\leq r \frac{s}{t-r+1}
\end{aligned}$$

The last step uses the following: For  $0 < x < 1$ ,  $1 - (1-x)^r < rx$ . This can be proved by showing that the line  $y = rx$  is tangent to  $g(x) = 1 - (1-x)^r$  at  $(0, 0)$  and that  $g(x)$  is concave down on the interval  $[0, 1]$ .  $\square$

**Corollary 1.7.** *In the proposition, when  $r < t/2$ , the upper bound can be simplified to  $2rs/t$ .*

## 2 The Noisy Coding Theorem

Let  $D$  be the Hamming distance in  $\mathbb{F}_2^n$ . For any nonnegative integer  $t$  and any  $y \in \mathbb{F}_2^n$ , let  $B_t(y) = \{y' \in \mathbb{F}_2^n : D(y', y) \leq t\}$ .

Let  $C \subseteq \mathbb{F}_2^n$  be a code and let  $M = |C|$ . We define the  $t$ -ball decoding strategy as follows. For a received vector  $y$ ,

- If  $B_t(y)$  contains a unique codeword  $x$  return  $x$ .
- Otherwise, return nothing.

Consider sending the vector  $x \in C$  and receiving the vector  $y \in \mathbb{F}_2^n$ . Given  $y$  the decoder returns  $x$  if

$$x \in B_t(y) \text{ and } \forall x' \in C \setminus \{x\}, x' \notin B_t(y)$$

This is equivalent to

$$y \in B_t(x) \text{ and } \forall x' \in C \setminus \{x\}, y \notin B_t(x')$$

Therefore the decoder fails to return  $x$  when either

1.  $y \notin B_t(x)$ , or
2.  $y \in B_t(x)$  and there exists some  $x' \in C$  such that  $x' \neq x$  and  $y \in B_t(x')$

**Definition 2.1.** For a given  $t$  and  $x \in F^n$  let

$$V_t(x) = B_t(x) \cap \left( \bigcup_{\substack{x' \in C \\ x' \neq x}} B_t(x') \right)$$

Notice that  $V_t(x)$  depends on the code  $C$ . We define two disjoint sets of  $x, y$  pairs corresponding to the two types of errors:

$$\begin{aligned} E_1 &= \{(x, y) : y \notin B_t(x)\} \\ E_2 &= \{(x, y) : y \in V_t(x)\} \\ E &= E_1 \cup E_2 \end{aligned}$$

Now for a fixed  $x$  the probability of a decoding error is

$$P_x(E) = P_x(E_1) + P_x(E_2)$$

The first term is easy to bound, but the second is more problematic and will require us to average over all codes.

For the first term, we assume the channel is symmetric and memoryless, so that  $P_x(y) = \alpha^w(1 - \alpha)^{n-w}$  where  $w = \text{wt}(x - y)$ . Then

$$P_x(E_1) = \sum_{w=t+1}^n \binom{n}{w} \alpha^w (1 - \alpha)^{n-w}$$

which is the probability that the sum of  $n$  iid Bernoulli random variables is larger than  $w$ . The expected number of errors is  $\alpha n$ . We will choose  $t$  a bit bigger than the expected number of errors  $t = \lceil (\alpha + \delta)n \rceil$  where  $\delta > 0$  will be determined later. Then by the law of large numbers

$$\begin{aligned} P_x(E_1) &\leq P(S_n > \lceil (\alpha + \delta)n \rceil) \\ &= P((S_n/n - \alpha) > \delta) \\ &< \frac{\alpha(1 - \alpha)}{\delta^2 n} \end{aligned}$$

This is independent of  $x$  so summing over  $x$  we will have

$$P(E_1) = \sum_{x \in \mathcal{C}} p(x) P_x(E_1) < \frac{\alpha(1 - \alpha)}{\delta^2 n}$$

As we make  $\delta$  (and therefore  $t$ ) larger this bound on  $P(E_1)$  diminishes, but  $P(E_2)$  increases. The noisy coding theorem will choose an appropriate  $\delta$  to bound both terms.

**Theorem 2.2 (Noisy Coding).** *Let  $\pi : \{0, 1\} \rightarrow \{0, 1\}$  be a binary symmetric memoryless channel, with crossover probability  $\alpha < 1/2$ . The capacity of the channel is  $\mathcal{K} = 1 - h(\alpha)$ . Let  $R < \mathcal{K}$  and let  $\epsilon > 0$  be given. Choose  $\delta$  such that  $R < 1 - h(\alpha + \delta) < \mathcal{K}$ . Let  $n$  be such that*

$$n > \begin{cases} \frac{2\alpha(1 - \alpha)}{\delta^2 \epsilon} \\ \frac{2 - \log(\epsilon)}{1 - h(\alpha + \delta) - R} \end{cases}$$

*Then there exists a code of length  $n$  and rate at least  $R$  for which the probability of error is less than  $\epsilon$ .*

PROOF: Let  $t = \lceil (\alpha + \delta)n \rceil$ . We will use the  $t$ -ball decoding strategy. We continue to use  $E_1$  and  $E_2$  to denote the two types of errors.

We showed above that, independent of the choice of code,

$$\begin{aligned} P(E_1) &< \frac{\alpha(1 - \alpha)}{\delta^2 n} \\ &\leq \epsilon/2 \end{aligned}$$

by our choice of  $n$ .

For  $E_2$  we consider all possible codes with  $M = \lceil 2^{nR} \rceil$ . Let

$$\mathcal{C} = \{C \subseteq \mathbb{F}_2^n : |C| = M\}$$

Clearly

$$|\mathcal{C}| = \binom{2^n}{M}$$

Give  $\mathcal{C}$  the uniform distribution and consider the channel map

$$\mathcal{C} \xrightarrow{\theta} \mathbb{F}_2^n$$

where

$$P_C(x) = \begin{cases} \frac{1}{M} & \text{if } x \in C \\ 0 & \text{otherwise} \end{cases}$$

We then have a Markov chain:

$$\mathcal{C} \longrightarrow \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^n \tag{1}$$

The distribution on  $\mathcal{C} \times \mathbb{F}_2^n \times \mathbb{F}_2^n$  determined by the Markov chain is

$$\begin{aligned} P(C, x, y) &= P(C)P_C(x)P_x(y) \\ &= \begin{cases} \frac{1}{\binom{2^n}{M}} \frac{1}{M} \alpha^w (1 - \alpha)^{n-w} & \text{if } x \in C \text{ where } D(x, y) = w \\ 0 & \text{if } x \notin C \end{cases} \end{aligned}$$

Let

$$E_2 = \{(C, x, y) : x \in B_t(y) \cap C \text{ and } |B_t(y) \cap C| \geq 2\}$$

We will show that  $P(E_2) < \epsilon/2$ . Then since

$$P(E_2) = \sum_{C \in \mathcal{C}} P(C)P_C(E_2) < \epsilon/2$$

There must be some particular  $C \in \mathcal{C}$  with  $P_C(E_2) < \epsilon/2$ . This establishes the theorem.

The distribution on  $\mathcal{C} \times \mathbb{F}_2^n \times \mathbb{F}_2^n$  is a Markov chain in the reverse direction to (1)

$$\mathcal{C} \longleftarrow \mathbb{F}_2^n \longleftarrow \mathbb{F}_2^n \tag{2}$$

It is easy to see that  $P(y) = 1/2^n$  and that  $P_y(x) = \alpha^w (1 - \alpha)^{n-w}$  where  $w = \text{wt}(y - x)$ . Since there are  $\binom{2^n - 1}{M - 1}$  codes containing  $x$ ,  $P_x(C) = \left(\binom{2^n - 1}{M - 1}\right)^{-1}$  for  $x \in C$  and  $P_x(C) = 0$  otherwise.

$$P(E_2) = \sum_{y \in \mathbb{F}_2^n} P(y) \sum_{x \in B_t(y)} P_y(x) \sum_{\substack{x \in C \\ |C \cap B_t(y)| \geq 2}} P_x(C)$$

It is clearly sufficient to show that the inner sum is always less than  $\epsilon/2$ . For fixed  $x$  and  $y$ , the inner sum is the probability of choosing a code  $C$ —from the  $\binom{2^n - 1}{M - 1}$  codes containing  $x$ —which contains at least one more element of  $B_t(y)$ . This is the proportion of ways to choose  $r = M - 1$  objects from  $t = 2^n - 1$  objects and get at least one from a set containing  $s = |B_t(y)| - 1$  objects, so Corollary 1.7, applies. By the tail inequality,

$$\begin{aligned} |B_t(y)| &= \sum_{k=0}^{\lfloor (\alpha + \delta)n \rfloor} \binom{n}{k} \\ &< 2^{nh(\alpha + \delta)} \end{aligned}$$

Therefore the inner sum is at most

$$2(M-1) \frac{2^{nh(\alpha+\delta)} - 1}{2^n - 1} \leq 2^{1+n(h(\alpha+\delta)+R-1)}$$

Since

$$\begin{aligned} n &> \frac{2 - \log(\epsilon)}{1 - h(\alpha + \delta) - R} \\ n(h(\alpha + \delta) + R - 1) &< \log(\epsilon) - 2 \end{aligned}$$

Thus

$$\begin{aligned} \sum_{\substack{x \in C \\ |C \cap B_t(y)| \geq 2}} P_x(C) &< 2^{1+\log \epsilon - 2} \\ &= \epsilon/2 \end{aligned}$$

□