Coding Theory
Lecture 5

Lecture by Dr. Elette Boyle
Typeset by Steven Karas

2017-11-30
Last edited 18:19:13 2017-11-30

Disclaimer These lecture notes are based on the lecture for the course Coding Theory, taught by Dr. Elette Boyle at IDC Herzliyah in the fall semester of 2017/2018. Sections may be based on the lecture slides written by Dr. Elette Boyle.

Recap

- Upper/Lower Bounds
  - Singleton Bound (neg)
  - Hamming Bound (neg)
  - GV Bound (pos)
  - Plotkin Bound (neg)

Agenda

- Stochastic Model of Errors
- Shannon's Model
- Commonly studied channels
- Capacity of a channel
- Capacity of Binary Symmetric Channel with parameter p

1 Shannon's Model of Errors

Historically, Shannon published "A Mathematical Theory of Communication" in 1948. Presented a framework of encoding messages over a channel which transmitted messages between two entities.

Noise Model Given an input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}$. Input messages are vectors of $x \in \mathcal{X}$, and output messages are vectors of $y \in \mathcal{Y}$. Channel is memoryless, meaning that the noise acts independently on each transmitted symbol. For our purposes, we will only consider discrete channels where $\mathcal{X}, \mathcal{Y}$ are finite.

The action of the channel is defined by a transition matrix:

$$\begin{pmatrix} \dots & \dots & \dots \\ \dots & \Pr(y|x) & \dots \\ \dots & \dots & \dots \end{pmatrix}$$

1.1 Error Correction in Stochastic Models

Hamming model is worst case (error < bound => correct decoding guaranteed).

Shannon (stochastic) model is that there is some distribution defined for the noise. As such, we cannot guarantee correct decoding. Our goal is for every message, to correctly decode with high probability. Ideally, we can define a code such that the decoding error as a function of n is $\lim_{n\to\infty} f(n) = 0$.

For some encoding algorithm E and a corresponding decoding algorithm D:

$$\Pr_{\text{channel transformation}}[D(E(\vec{m})) \neq \vec{m}]$$

A central challenge for us, as before, is to understand optimal tradeoffs between the rate of a code and "how much error" it can decode from. But how do we measure "how much error" for a given channel? The notion of capacity of a channel is a measure of how high of a rate of information conveyance you can possibly achieve across it.

1.2 Channel Capacity

The capacity of a channel is the value $C \in \mathbb{R}$ such that:

For any $R < C$, there exists some code E, D with rate R and negligible decoding error ($\lim_{n\to\infty} f(n) = 0$). For any $R > C$, then for every code E, D, the decoding error probability is bounded below by some constant.

Note, that the best possible capacity is 0, and the best possible capacity is 0.

Notation We write $\vec{e} \sim \mathrm{BSC}_p$ to denote the error pattern sampled from BSC_p.

That is, each position of the vector $\vec{e}$ is independently chosen as 1 with probability p and 0 with probability $1 - p$.

the rough takeaway here is that the capacity is $1 - H(p)$

Shannon's Capacity Theorem for $\mathbf{BSC}_p$ For every $p, \varepsilon \in \mathbb{R}$ such that $0 \leq p \leq \frac{1}{2}$ and $0 \leq \varepsilon \leq \frac{1}{2} - p$, the following holds for sufficiently large n:

1. There exists $\delta > 0$, an encoding function $E : \{0,1\}^k \to \{0,1\}^n$, and a decoding function $D : \{0,1\}^n \to \{0,1\}^k$ such that $k \leq \lfloor (1 - H(p + \varepsilon))n \rfloor$. For any $\vec{m} \in \{0,1\}^k$:

$$\Pr_{\vec{e}\sim\mathrm{BSC}_p}[D(E(\vec{m}) + \vec{e}) \neq \vec{m}] \leq 2^{-\delta n}$$

2. if $k \geq \lceil (1 - H(p) + \varepsilon) \rceil n$, then for any $E : \{0,1\}^k \to \{0,1\}^n$, and a decoding function $D : \{0,1\}^n \to \{0,1\}^k$ then there exists a message $\vec{m} \in \{0,1\}^k$ such that:

$$\Pr_{\vec{e}\sim\mathrm{BSC}_p}[D(E(\vec{m}) + \vec{e}) \neq \vec{m}] \geq \frac{1}{2}$$

- If $p = 0$, then $\forall \varepsilon > 0$, $1 - H(p) + \varepsilon > 1$, so claim holds directly. The alternative would be encoding n bits of information with fewer than n bits.
- Take $0 < p \leq \frac{1}{2}$. Assume (for contradiction) that there exists some code E, D with $k \geq \lceil (1 - H(p) + \varepsilon) \rceil n$ such that $\forall \vec{m} \in \mathcal{X}$:

$$\Pr_{\vec{e}\sim\mathrm{BSC}_p}[D(E(\vec{m}) + \vec{e}) \neq \vec{m}] < \frac{1}{2}$$

Sketch: Show that the decoding domains are large and disjoint. Then by packing, argue that we have too many such domains, which implies a bound on k

For each $\vec{m} \in \mathcal{X}$, define $D_{\vec{m}}$ to be the set of received words $\vec{y} \in \mathcal{Y}$ such that D decodes $\vec{y}$ to $\vec{m}$.

For some $\vec{m}$, define a shell with width $2\gamma > 0$ around $E(\vec{m})$ as:

$$S_{\vec{m}} = B(E(\vec{m}), (1+\gamma)pn) \setminus B(E(\vec{m}), (1-\gamma)pn)$$

$$\Pr_{\vec{e}\sim\mathrm{BSC}_p}[E(\vec{m}) + \vec{e} \notin D_{\vec{m}}] < \frac{1}{2}$$

$$\Pr_{\vec{e}\sim\mathrm{BSC}_p}[E(\vec{m}) + \vec{e} \notin S_{\vec{m}}] < 2^{-\Omega(\gamma^2 n)}$$

By the Chernoff bound, note that $\Pr[|\vec{e}| > (1+\gamma)\mu \ldots] \le e^{\gamma^2 \mu \cdots}$.

$$\Pr_{\vec{e} \sim \mathrm{BSC}_p}[E(\vec{m}) + \vec{e} \notin D_{\vec{m}} \cap S_{\vec{m}}] < \frac{1}{2} + 2^{-\Omega(\gamma^2 n)}$$

$$\Pr_{\vec{e} \sim \mathrm{BSC}_p}[E(\vec{m}) + \vec{e} \in D_{\vec{m}} \cap S_{\vec{m}}] > \frac{1}{2} - 2^{-\Omega(\gamma^2 n)} \ge \frac{1}{4} \text{ for sufficiently large } n$$

We want to lower bound $|D_{\vec{m}} \cap S_{\vec{m}}|$ (in order to lower bound $D_{\vec{m}}$).

Sketch: if we can argue that no individual point in $D_{\vec{m}} \cap S_{\vec{m}}$ has too large probability mass on its own, then it must be that $D_{\vec{m}} \cap S_{\vec{m}}$ has many $\vec{y}$s to reach total probability mass $> 1/4$

$$\Pr_{\vec{e} \sim \mathrm{BSC}_p}[E(\vec{m}) + \vec{e} \in D_{\vec{m}} \cap S_{\vec{m}}] \le |D_{\vec{m}} \cap S_{\vec{m}}| \cdot p_{\max}$$

where $p_{\max} = \max_{\vec{y} \in D_{\vec{m}} \cap S_{\vec{m}}} \Pr[E(\vec{m}) + \vec{e} = \vec{y}]$.
Note that we can place a bound on $p_{\max}$ because of the definition of $\vec{e}$ and d.

$$\underbrace{(1-p)^{n-d}}_{\text{flip 0s}} \cdot \underbrace{(p)^d}_{\text{flip 1s}}$$

For $p \le \frac{1}{2}$, this function monotonically decreases with d, so the point with highest probability lies on the inner edge of the shell, with distance $(1-\gamma)pn$.

$$\begin{aligned}
p_{\max} &\le p^{(1-\gamma)pn}(1-p)^{n-(1-\gamma)pn} \\
&= \left(\frac{1-p}{p}\right)^{\gamma pn} p^{pn}(1-p)^{(1-p)n} \\
&= \left(\frac{1-p}{p}\right)^{\gamma pn} 2^{-nH(p)} \\
H(p) &= -p\log p - (1-p)\log(1-p)
\end{aligned}$$

Combining with before, we get that:

$$|D_{\vec{m}} \cap S_{\vec{m}}| \cdot p_{\max} \ge \Pr_{\vec{e} \sim \mathrm{BSC}_p}[E(\vec{m}) + \vec{e} \in D_{\vec{m}} \cap S_{\vec{m}}] \ge \frac{1}{4}$$

$$|D_{\vec{m}} \cap S_{\vec{m}}| \ge \frac{1}{4}\frac{1}{p_{\max}} \ge \frac{1}{4}\left(\frac{1-p}{p}\right)^{-\gamma pn} 2^{nH(p)}$$

Recall that $|D_{\vec{m}}| \ge |D_{\vec{m}} \cap S_{\vec{m}}|$, and that $D_{\vec{m}}$ form a fully covering partition over $\{0,1\}^n$.
Take $\gamma = \frac{1}{2}\frac{\varepsilon}{p\log\left(\frac{1-p}{p}\right)}$, recalling that γ is constant, and $\gamma \ge 0$ for all $p < \frac{1}{2}$.

$$\begin{aligned}
2^n = |\{0,1\}^n| &= \sum_{\vec{m} \in \{0,1\}^k} |D_{\vec{m}}| \\
&\ge \sum_{\vec{m} \in \{0,1\}^k} |D_{\vec{m}} \cap S_{\vec{m}}| \\
&\ge \sum_{\vec{m} \in \{0,1\}^k} \frac{1}{4}\left(\frac{1-p}{p}\right)^{-\gamma pn} 2^{nH(p)} \\
&= 2^k \cdot 2^{-2} \cdot 2^{-\gamma pn \log(\ldots)} \cdot 2^{H(p)n} \\
&= 2^{k + H(p)n - 2 - \varepsilon n/2} \\
&> 2^{k + H(p)n - \varepsilon n}
\end{aligned}$$

Which for sufficiently large n means that $n > k + H(p)n - \varepsilon n$, which implies that $k < (1 - H(p) + \varepsilon)n$.
This contradicts our original assumption on k.

This implies that the capacity of BSC_p is $1 - H(p)$

Entropy function The entropy function $H_q(p)$

1.3 Example: The Binary Symmetric Channel

Let $0 \leq p \leq 1$. The binary symmetric channel BSC_p with crossover probability p is defined as:

$$\mathcal{X} = \mathcal{Y} = \{0,1\}$$

The transition matrix is:

$$M = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

That is that each bit is flipped independently with probability p.

Capacity

$$\begin{aligned} p = 0 &\longrightarrow C = 1 \\ p = 1/2 &\longrightarrow C = 0 \\ p = 1 &\longrightarrow C = 1 \end{aligned}$$

1.4 Example: q-ary Symmetric Channel

This is the generalization of the BSC_p to larger alphabets.

$$\mathcal{X} = \mathcal{Y} = [q]$$

$$\forall x, y \in [q],\ M_{xy} = \begin{cases} 1-p & \text{if } y = x \\ \frac{p}{q-1} & \text{else} \end{cases}$$

1.5 Example: Binary Erasure Channel

Let $0 \leq \alpha \leq 1$.

$$\mathcal{X} = \{0,1\}$$
$$\mathcal{Y} = \{0,1,?\}$$
$$M = \begin{pmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1-\alpha & \alpha \end{pmatrix}$$

1.6 Example: Additive Gaussian White Noise (AGWN) Channel

Note that while we only consider discrete channels for our purposes, continuous channels are very useful.

$$\mathcal{X} = \{0,1\}$$
$$\mathcal{Y} = \mathbb{R}$$