Lecture 18: Shanon's Channel Coding Theorem

Channel

Definition (Channel)

A channel is defined by $\Lambda = (X, Y, \Pi)$, where X is the set of input alphabets, Y is the set of output alphabets and Π is the transition probability of obtaining a symbol $y \in Y$ if the input symbol is $x \in X$.

- For example: A Binary Symmetric Channel with flipping probability p (i.e., p-BSC) is a channel with $X = \{0, 1\}$ and $Y = \{0, 1\}$, and the probability of obtaining b given input symbol b is $(1 - p)$ and the probability of obtaining $(1 - b)$ given input symbol b is p

Channel Capacity

Definition (Capacity)

The capacity of a channel is defined by:

$$C(\Lambda) = \max_{\text{Dist } p \text{ over } X} H(Y) - H(Y|X)$$

- Note that it is not necessary that the maximization happens when p is the uniform distribution over X
- For p-BSC, the maximization happens for $p = U_X$ and the capacity is $1 - h(p)$

Shannon's Channel Coding Theorem

Theorem (Shanon's Channel Coding Theorem)

For every channel Λ, there exists a constant $C = C(\Lambda)$, such that for all $0 \leqslant R < C$, there exists n_0, such that for all $n \geqslant n_0$, there exists encoding and decoding algorithms Enc *and* Dec *such that:*

- Enc: $\{1, \ldots, M = 2^{Rn}\} \to X^n$, *and*
- $\Pr[\mathrm{Dec}(\Pi(\mathrm{Enc}(m))) = m] \geqslant 1 - \exp(-\Omega(n))$

- English Version: For every channel, there exists a constant capacity, such that for all rate less than the capacity, (for large enough n), we can reliably push information across the channel at that rate

Coding Theorem for BSC

Let $Z(n, p)$ be n independent trials of a Bernoulli variable with probability of heads being p

Theorem

For all p, there exists $C = 1 - h(p)$, such that for all $0 \leqslant R = 1 - h(p) - \varepsilon$ and $\varepsilon > 0$, there exists n_0, such that for all $n \geqslant n_0$, there exists encoding and decoding algorithms Enc *and* Dec *such that:*

- Enc: $\{0, 1\}^{Rn} \to \{0, 1\}^n$, *and*
- $\Pr_{z \sim Z(n,p)}[\mathrm{Dec}(\mathrm{Enc}(m) + z) = m] \geqslant 1 - \exp(-\Omega(n))$

- In fact, there exists a binary linear code that achieves this rate
- Further, a random binary linear code achieves this rate with probability exponentially close to 1

Proof of Coding Theorem for BSC

- Define $k = (1 - h(p) - \varepsilon)n$ and $\ell = k + 1$
- We shall show that there exists an encoding scheme Enc^* using probabilistic methods
- Let $\mathrm{Enc} \colon \{0,1\}^\ell \to \{0,1\}^n$ be a random map
- Let $\mathrm{Dec}(y)$ be the maximum likelihood decoding, i.e. it decodes y to the nearest codeword
- Fix a message $m \in \{0,1\}^\ell$
- We are interested in: Expected (over random Enc) decoding error probability

$$\mathrm{err}(m) := \mathop{\mathbb{E}}_{\mathrm{Enc}}\left[\Pr_{z \sim Z(n,p)}[\mathrm{Dec}(\mathrm{Enc}(m) + z) \neq m]\right]$$

- Note that, we have:

$$\mathrm{err}(m) \leqslant \mathop{\mathbb{E}}_{\mathrm{Enc}}\left[\Pr_{z \sim Z(n,p)}[\mathrm{wt}(z) \geqslant (p+\varepsilon)n] \right.$$
$$\left. + \Pr_{z \sim Z(n,p)}[\mathrm{wt}(z) \leqslant (p+\varepsilon)n \ \wedge \ \mathrm{Dec}(\mathrm{Enc}(m) + z) \neq m]\right]$$

Proof of Coding Theorem for BSC (continued)

- First error term is at most $\exp(-\Omega(\varepsilon^2 n))$ by Chernoff Bound
- Let $p(z)$ be the probability of $z \sim Z(n,p)$
- $\mathbf{1}(E)$ represents the indicator variable for the event E
- So, we have: $\mathrm{err}(m) \leqslant \mathbb{E}_{\mathrm{Enc}}\left[\mathrm{err}_1 + \mathrm{err}_2(m, \mathrm{Enc})\right]$, where:

$$\mathrm{err}_1 = \exp(-\Omega(\varepsilon^2 n))$$

$$\mathrm{err}_2(m, \mathrm{Enc}) = \sum_{z \in \mathrm{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot \mathbf{1}(\mathrm{Dec}(\mathrm{Enc}(m) + z) \neq m)$$

- By linearity of expectation, we get:
$\mathrm{err}(m) \leqslant \exp(-\Omega(\varepsilon^2 n)) + \mathbb{E}_{\mathrm{Enc}}\left[\mathrm{err}_2(m, \mathrm{Enc})\right]$

Proof of Coding Theorem for BSC (continued)

- We need to bound only: $\mathbb{E}_{\mathsf{Enc}}\left[\mathrm{err}_2(m, \mathsf{Enc})\right]$, which turns out to be (by swapping $\sum$ and $\mathbb{E}$ operators):

$$\sum_{z \in \mathsf{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot \mathop{\mathbb{E}}_{\mathsf{Enc}}\left[\mathbf{1}(\mathsf{Dec}(\mathsf{Enc}(m)+z) \neq m)\right]$$

$$= \sum_{z \in \mathsf{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot \mathop{\mathbb{E}}_{\mathsf{Enc}}\left[\exists m' \neq m \colon \mathbf{1}(\mathsf{Dec}(\mathsf{Enc}(m)+z) = m')\right]$$

$$= \sum_{z \in \mathsf{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot \Pr_{\mathsf{Enc}}\left[\exists m' \neq m \colon \mathsf{Dec}(\mathsf{Enc}(m)+z) = m'\right]$$

$$\leqslant \sum_{z \in \mathsf{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot \Pr_{\mathsf{Enc}}\left[\exists m' \neq m \colon c' \in c + \mathsf{Ball}_2(n,(p+\varepsilon)n)\right]$$

Here c and c', respectively, are $\mathsf{Enc}(m)$ and $\mathsf{Enc}(m')$

Proof of Coding Theorem for BSC (continued)

- By union bound, we get:

$$\leqslant \sum_{z \in \mathsf{Ball}_2(n,(p+\varepsilon)n)} p(z) \sum_{m'\colon m' \neq m} \Pr_{\mathsf{Enc}}\left[c' \in c + \mathsf{Ball}_2(n,(p+\varepsilon)n)\right]$$

$$\leqslant \sum_{z \in \mathsf{Ball}_2(n,(p+\varepsilon)n)} p(z) \sum_{m'\colon m' \neq m} \frac{\mathsf{Vol}_2(n,(p+\varepsilon)n)}{2^n}$$

$$\leqslant \sum_{z \in \mathsf{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot 2^{\ell} \cdot \frac{2^{h(p)n}}{2^n} = \sum_{z \in \mathsf{Ball}_2(n,(p+\varepsilon)n)} p(z) 2 \cdot 2^{-\varepsilon n}$$

$$= 2 \cdot 2^{-\varepsilon n} = \exp(-\Omega(n))$$

- Overall, we get: For a fixed m, the expected decoding error (over a randomly chosen encoding function) is $\mathsf{err}(m) \leqslant \mathbb{E}_{\mathsf{Enc}}\left[\mathsf{err}_1 + \mathsf{err}_2(m, \mathsf{Enc})\right] \leqslant \exp(-\Omega(n))$

Proof of Coding Theorem for BSC (continued)

- Therefore,

$$\mathop{\mathbb{E}}_{\mathrm{Enc}}\left[\mathop{\mathbb{E}}_{m \xleftarrow{\$} \{1,\dots,2^\ell\}}\left[\Pr_{z\sim Z(n,p)}[\mathrm{Dec}(\mathrm{Enc}(m)+z)\neq m]\right]\right]$$

$$= \mathop{\mathbb{E}}_{m \xleftarrow{\$} \{1,\dots,2^\ell\}}\left[\mathop{\mathbb{E}}_{\mathrm{Enc}}\left[\Pr_{z\sim Z(n,p)}[\mathrm{Dec}(\mathrm{Enc}(m)+z)\neq m]\right]\right]$$

$$\leqslant \mathop{\mathbb{E}}_{m \xleftarrow{\$} \{1,\dots,2^\ell\}}[\exp(-\Omega(n))] = \exp(-\Omega(n))$$

- So, there exists an Enc^* such that the expected (over random messages) decoding error probability is at most $\exp(-\Omega(n))$
- By pigeon hole principle, for this choice of Enc^*, at most half the messages in $\{1,\dots,2^\ell\}$ have decoding error probability $\geqslant 2\cdot\exp(-\Omega(n))$
- So, for this choice of Enc^* there exists a subset of $\{1,\dots,2^\ell\}$ of size 2^k such that each message has decoding error probability $\leqslant 2\cdot\exp(-\Omega(n)) = \exp(-\Omega(n))$

Additional Comments

- If we show that a random linear encoding Enc succeeds then we do not need to perform an averaging over m, because the decoding error probability for a particular m is identical to the decoding error probability for *any* m (because, in linear codes, "the view from a codeword c about the universe is identical to the view of any codeword c about the universe")
- We can also show that for $1 - \exp(-\Omega(n))$ fraction of Enc the decoding error is exponentially small (because, decoding error is bounded by 1 and we can perform an averaging argument)

Intuitively,

Theorem (Converse of Shannon's Channel Coding Theorem)

For any channel Λ, there exists $C = C(\Lambda)$, such that for all $R > C$, and any encoding and decoding functions Enc *and* Dec*, respectively, (for all n) the decoding error is at least a constant when $m \xleftarrow{\$} \{1, \dots, 2^{Rn}\}$*