Lecture 18: Shannon's Channel Coding Theorem
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 $\Pi$ 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 $p$. 
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 \( 1 - h(p) \).
Theorem (Shannon’s Channel Coding Theorem)

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

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

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 \( 0 \leq R = 1 - h(p) - \varepsilon \) and \( \varepsilon > 0 \), there exists \( n_0 \), such that for all \( n \geq n_0 \), there exists encoding and decoding algorithms \( \text{Enc} \) and \( \text{Dec} \) such that:

- \( \text{Enc}: \{0, 1\}^{Rn} \rightarrow \{0, 1\}^n \), and
- \( \Pr_{z \sim Z(n,p)}[\text{Dec}(\text{Enc}(m) + z) = m] \geq 1 - \exp(-\Omega(n)) \)

- 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 \( \text{Enc}^* \) using probabilistic methods
- Let \( \text{Enc}: \{0, 1\}^\ell \rightarrow \{0, 1\}^n \) be a random map
- Let \( \text{Dec}(y) \) 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 \( \text{Enc} \)) decoding error probability

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

Note that, we have:

\[
\text{err}(m) \leq \mathbb{E}_{\text{Enc}} \left[ \Pr_{z \sim Z(n, p)} [\text{wt}(z) \geq (p + \varepsilon)n] \right] + \mathbb{E}_{\text{Enc}} \left[ \Pr_{z \sim Z(n, p)} [\text{wt}(z) \leq (p + \varepsilon)n \land \text{Dec}(\text{Enc}(m) + z) \neq m] \right]
\]
First error term is at most $\exp(-\Omega(\varepsilon^2 n))$ by Chernoff Bound

Let $p(z)$ be the probability of $z \sim Z(n, p)$

$1(E)$ represents the indicator variable for the event $E$

So, we have: $\text{err}(m) \leq \mathbb{E}_{\text{Enc}}[\text{err}_1 + \text{err}_2(m, \text{Enc})]$, where:

\[
\text{err}_1 = \exp(-\Omega(\varepsilon^2 n))
\]

\[
\text{err}_2(m, \text{Enc}) = \sum_{z \in \text{Ball}_2(n, (p+\varepsilon)n)} p(z) \cdot 1(\text{Dec}(\text{Enc}(m) + z) \neq m)
\]

By linearity of expectation, we get:

$\text{err}(m) \leq \exp(-\Omega(\varepsilon^2 n)) + \mathbb{E}_{\text{Enc}}[\text{err}_2(m, \text{Enc})]$
We need to bound only: $\mathbb{E}_{Enc} [err_2(m, Enc)]$, which turns out to be (by swapping $\sum$ and $\mathbb{E}$ operators):

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

$$
= \sum_{z \in \text{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot \mathbb{E}_{Enc} [\exists m' \neq m: 1(\text{Dec}(Enc(m) + z) = m')] 
$$

$$
= \sum_{z \in \text{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot \text{Pr}_{Enc} [\exists m' \neq m: \text{Dec}(Enc(m) + z) = m'] 
$$

$$
\leq \sum_{z \in \text{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot \text{Pr}_{Enc} [\exists m' \neq m: c' \in c + \text{Ball}_2(n,(p+\varepsilon)n)] 
$$

Here $c$ and $c'$, respectively, are $\text{Enc}(m)$ and $\text{Enc}(m')$
By union bound, we get:

\[
\leq \sum_{z \in \text{Ball}_2(n,(p+\varepsilon)n)} p(z) \sum_{m' : m' \neq m} \Pr_{\text{Enc}}[c' \in c + \text{Ball}_2(n,(p+\varepsilon)n)]
\]

\[
\leq \sum_{z \in \text{Ball}_2(n,(p+\varepsilon)n)} p(z) \sum_{m' : m' \neq m} \frac{\text{Vol}_2(n,(p+\varepsilon)n)}{2^n}
\]

\[
\leq \sum_{z \in \text{Ball}_2(n,(p+\varepsilon)n)} p(z) \cdot 2^{h(p)n} \cdot \frac{2^{h(p)n}}{2^n} = \sum_{z \in \text{Ball}_2(n,(p+\varepsilon)n)} p(z) 2 \cdot 2^{-\varepsilon n}
\]

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

Overall, we get: For a fixed \( m \), the expected decoding error (over a randomly chosen encoding function) is

\[
\text{err}(m) \leq \mathbb{E}_{\text{Enc}}[\text{err}_1 + \text{err}_2(m,\text{Enc})] \leq \exp(-\Omega(n))
\]
Proof of Coding Theorem for BSC (continued)

Therefore,

\[
\mathbb{E}\left[ \mathbb{E}_{m \leftarrow \{1, \ldots, 2^\ell\}} \left[ \Pr_{z \sim Z(n, p)} \left[ \text{Dec}(Enc(m) + z) \neq m \right] \right] \right]
\]

\[
= \mathbb{E}_{m \leftarrow \{1, \ldots, 2^\ell\}} \left[ \mathbb{E}_{Enc} \left[ \Pr_{z \sim Z(n, p)} \left[ \text{Dec}(Enc(m) + z) \neq m \right] \right] \right]
\]

\[
\leq \mathbb{E}_{m \leftarrow \{1, \ldots, 2^\ell\}} \left[ \exp(-\Omega(n)) \right] = \exp(-\Omega(n))
\]

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 \(\{1, \ldots, 2^\ell\}\) have decoding error probability \(\geq 2 \cdot \exp(-\Omega(n))\).

So, for this choice of \(Enc^*\) there exists a subset of \(\{1, \ldots, 2^\ell\}\) of size \(2^k\) such that each message has decoding error probability \(\leq 2 \cdot \exp(-\Omega(n)) = \exp(-\Omega(n))\).
If we show that a random linear encoding $\text{Enc}$ 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 $\text{Enc}$ 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 $\Lambda$, 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 $m \leftarrow \{1, \ldots, 2^{Rn}\}$. 