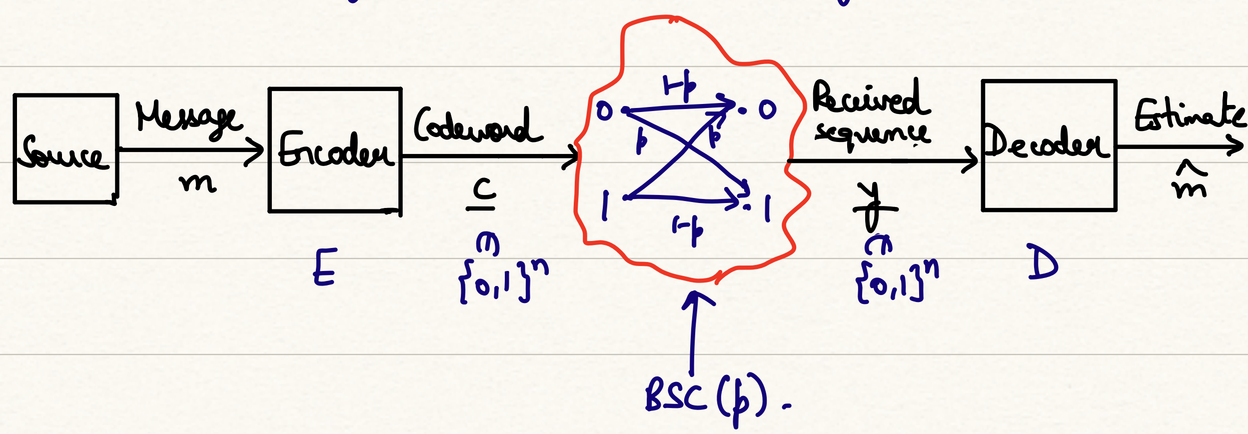
Lecture 5: Shannon's coding theorem

In this lecture, we shall take a look at an important result in error-control coding / information theory that pins down the tradeoff between the rate of a code and its error-correcting capability, in the setting of stochastic / random noise.

We work in the setting of transmission of binary codewords over a BSC(p).



WLOG, we assume that $0 < p < \frac{1}{2}$. The case where $p = 0$ is trivial and we note that if $p > \frac{1}{2}$, we can convert the channel to BSC(p'), where $p' = 1 - p < \frac{1}{2}$, by simply flipping all bits in $\underline{y}$.

Let the message set be $\mathcal{M}$, with $|\mathcal{M}| = 2^{nR}$, where R is the code rate.

The task at hand is to design (E, D) so as to guarantee "reliable reconstruction/recovery" of the message, i.e.,

$$\Pr[\hat{m} \neq m] \xrightarrow[n\to\infty]{} 0.$$

↳ over randomness in encoding, channel noise, & decoding.

<u>Theorem</u> (Shannon's noisy channel coding theorem). Fix $p \in (0, \frac{1}{2})$. There exists a real number $C = C(p)$ such that

(i) For any rate $R < C(p)$, there exist (sequences of) encoding/decoding functions $\{(E_n, D_n)\}_{n \geq 1}$ s.t. $\Pr[\hat{m} \neq m] \xrightarrow[n\to\infty]{} 0$.

["Achievability"]

(ii) For any rate $R > C(p)$, for any (sequences of) encoding/decoding functions $\{(E_n, D_n)\}_{n\geq 1}$, we have that $\exists$ message m s.t.

$$\Pr[\hat{m} = m] \xrightarrow[n\to\infty]{} 0.$$

["Converse"]

<u>Remark</u>: (i) Shannon's theorem above applies to the broader class of discrete memoryless channels (DMCs) of which the BSC is a part.

(ii) The quantity $C(p)$ is called the "capacity" of the BSC, since it represents a threshold, at rates below which reliable communication is possible, and at rates above which reliable communication is impossible.

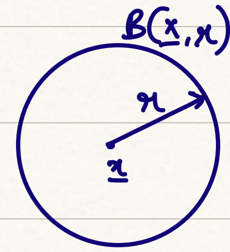
For the BSC, the capacity is

$$C(p) = 1 - h_b(p) \equiv 1 - h(p),$$

where $h(\cdot)$ is the binary entropy function:

$$h(p) = -p\log_2 p - (1-p)\log_2(1-p),$$

with $h(0) = h(1) \triangleq 0$.



First, a lemma. Let $B(\underline{x}, r)$ be the Hamming ball of radius r around $\underline{x} \in \{0,1\}^n$.

<u>Lemma</u>: We have that for any $\underline{x} \in \{0,1\}^n$,

$$2^{n(h(p)-o(1))} \leq \mathrm{Vol}(B(\underline{x}, np)) = \sum_{j=0}^{np} \binom{n}{j} \leq 2^{nh(p)}.$$

<u>Proof</u>: Note that

$$\frac{\mathrm{Vol}(B(\underline{x}, np))}{2^{nh(p)}} = \frac{\mathrm{Vol}(B(\underline{0}, np))}{2^{nh(p)}}$$

$$= \sum_{j=0}^{np} \binom{n}{j} \cdot p^{np} (1-p)^{n(1-p)}$$

$$\leq \sum_{j=0}^{np} \binom{n}{j} p^j (1-p)^{n-j}$$

$$= (1-p)^n \cdot \sum_{j=0}^{np} \binom{n}{j} \left(\frac{p}{1-p}\right)^j$$

$$\leq (1-p)^n \cdot \sum_{j=0}^{n} \binom{n}{j} \left(\frac{p}{1-p}\right)^j = (1-p)^n \cdot \left(1+\frac{p}{1-p}\right)^n = 1.$$

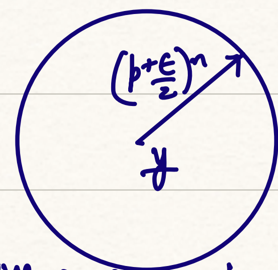
The proof of the lower bound follows via Stirling's inequality. $\boxtimes$

<u>Proof of achievability in Shannon's theorem</u>: Fix $R = 1 - h(p+\epsilon)$, for $\epsilon \in (0, \frac{1}{2} - p)$.

We induce a distribution over (E, D) and show that reliable reconst^n holds "with high probability". Hence, there <u>must exist</u> a deterministic (E, D) such that this property holds.

E: Pick $E(m) \sim \text{Unif}(\{0,1\}^n)$.

D: Given $y \in \{0,1\}^n$,



$$D(y) = \begin{cases} \bar{m}, & \text{if } \bar{m} \text{ is the unique } m \in \mathcal{M} \text{ s.t. } d(E(m), y) < (p+\frac{\epsilon}{2})n \\ \perp, & \text{o.w.} \end{cases}$$

Note that $y = E(m) + \underline{e}$, where $\underline{e} \sim (\text{Ber}(p))^{\otimes n}$.

We define the following "bad events":

$$\mathcal{E}_1 : \quad \sum_{i=1}^{n} e_i > \left(p + \frac{\epsilon}{2}\right)n$$

$$\mathcal{E}_2 : \quad \exists\, m' \neq m \text{ s.t. } d(E(m'), y) < \left(p + \frac{\epsilon}{2}\right)n.$$

We wish to bound $\Pr[\mathcal{E}_1 \cup \mathcal{E}_2] \leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2]$. Note that

$$\Pr[\mathcal{E}_1] \leq \begin{cases} e^{-c\varepsilon^2 n} & \text{(Chernoff bound)} \\ \frac{c'}{n} & \text{(Chebyshev inequality)} \end{cases}$$

In any case, $\Pr[\mathcal{E}_1] \xrightarrow[n\to\infty]{} 0$.

Further,

$$\Pr[\mathcal{E}_2] = \mathbb{E}\left[\Pr[\mathcal{E}_2 \mid m, E(m), \underline{e}]\right]$$

$$\leq \mathbb{E}\Big[\sum_{m'\neq m} \underbrace{\Pr\big[d(E(m'), \underline{y}) < (p+\tfrac{\epsilon}{2})n \mid m, E(m), \underline{e}\big]}\Big]$$

$$= \mathbb{E}\Big[\sum_{m'\neq m} \Pr\big[\text{unif. random length-}n\text{ string lies in a ball of radius } (p+\tfrac{\epsilon}{2})n\big]\Big]$$

$$= \sum_{m'\neq m} \frac{\mathrm{Vol}(B(\underline{0}, (p+\frac{\epsilon}{2})n))}{2^n}$$

$$\leq |\mathcal{M}| \cdot 2^{-n(1-h(p+\epsilon/2))} = 2^{n\cdot(h(p+\epsilon/2) - h(p+\epsilon))}$$

$$\xrightarrow[n\to\infty]{} 0 \qquad \blacksquare$$

Take-away: There exist "good" codes that guarantee reliable recovery in the setting of stochastic noise. But which codes are good?