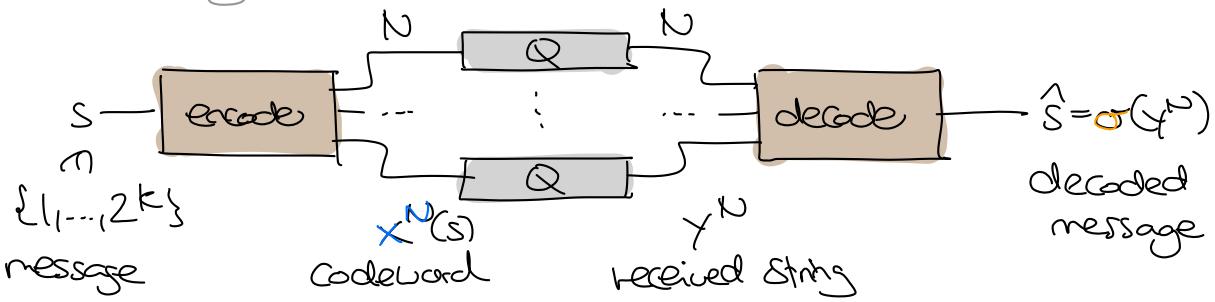
Proof of the Noisy Coding Theorem (§10)

Recall from Monday:



(N,K)-block code: $x^N: \{1,2,\dots,2^K\} \to \mathcal{X}^N$

Decoder: $\sigma: \mathcal{Y}^N \to \{\perp, 1,\dots,2^K\}$

Figures of merit:

* **rate:** $R := \frac{K}{N}$ bits per channel use
* **average prob. of (block) error** for uniform $S \in \{1,\dots,2^K\}$:

$$P_B = \Pr(\hat{S} \neq S) = \frac{1}{2^K} \sum_{s=1}^{2^K} \sum_{\hat{s} \neq s} P(\hat{s}|s)$$

Similarly for general $P(s)$

* **maximal probability of (block) error:**

$$P_{BM} = \max_s \Pr(\hat{S} \neq S \,|\, S=s) = \max_s \sum_{\hat{s} \neq s} P(\hat{s}|s) \geq P_B$$

enough to prove for P_B

Shannon's noisy coding theorem: Let $Q(y|x)$ channel.

Ⓐ **Achievability:**

If $\tilde{R} < C_1(Q)$: $\forall \delta > 0$: $\exists N_0$ $\forall N \geq N_0$: $\exists$ code with $\frac{K}{N} \geq \tilde{R}$ & $P_{BM} \leq \delta$

Ⓑ **Converse:**

If $\tilde{R} > C_1(Q)$: $\exists \delta > 0$ $\exists N_0$ $\forall N \geq N_0$: $\nexists$ code with $\frac{K}{N} \geq \tilde{R}$ & $P_B \leq \delta$

"weak converse" (also true $\forall \delta$ but will not prove this)

Proof of Achievability (Ⓐ)

Main tool: Jointly typical set for $P(x,y)$:

$$J_{N,\varepsilon}(P) = \left\{ (x^N, y^N) \text{ s.th. } x^N \in T_{N,\varepsilon}(P_x),\ y^N \in T_{N,\varepsilon}(P_y) \text{ and } (x^N, y^N) \in T_{N,\varepsilon}(P_{xy}) \right\}$$

Properties:

⓪ For all $(x^N, y^N) \in J_{N,\varepsilon}$: $2^{-N(H(X)+\varepsilon)} \le P(x^N) \le 2^{-N(H(X)-\varepsilon)}$ etc

① $\#J_{N,\varepsilon} \le 2^{N(H(X,Y)+\varepsilon)}$

② If $(X^N, Y^N) \overset{\text{IID}}{\sim} P(x,y)$: ← $(X_i, Y_i) \sim P$

$\Pr((X^N, Y^N) \in J_{N,\varepsilon}) \to 1$ as $N \to \infty$

③ If $\tilde{X}^N \overset{\text{IID}}{\sim} P(x)$ & $\tilde{Y}^N \overset{\text{IID}}{\sim} P(y)$ independent: ← $\tilde{X}_i, \tilde{Y}_i$ independent

$\Pr((\tilde{X}^N, \tilde{Y}^N) \in J_{N,\varepsilon}) \le 2^{-N(I(X:Y)-3\varepsilon)}$

Pf: LHS $\overset{\text{independence}}{=} \sum_{(x^N,y^N)\in J_{N,\varepsilon}} P(x^N)P(y^N) \overset{⓪}{\le} \#J_{N,\varepsilon} \cdot 2^{-N(H(X)-\varepsilon)} 2^{-N(H(Y)-\varepsilon)} \overset{①}{\le} 2^{-N(I(X:Y)-3\varepsilon)}$ □

Enough to prove: For all $P(x)$, $\tilde{R} < I(X:Y)$ w.r.t. $P(x,y) = P(x)Q(y|x)$, $\delta > 0$: ∃ sequence of (N, K)-block codes (one for each N) with $\frac{K}{N} \ge \tilde{R}$ s.th. $P_B \xrightarrow{N\to\infty} 0$

($K = K(N)$)

(Can always upgrade to P_{BT} via expurgation w/o changing rate much (→ last time))

Key idea: Choose code at random!

Random code: Let $K = \lceil N\tilde{R} \rceil$ and choose 2^K codewords at random:

$$X^N(1) = X_1(1) \quad X_2(1) \cdots X_N(1)$$
$$\vdots \qquad\quad \vdots \qquad\quad \vdots$$
$$X^N(2^K) = X_1(2^K) \; X_2(2^K) \cdots X_N(2^K)$$

IID $\sim P(x)$ — codeword by codeword, letter by letter

Lo (N,K)-code with $\frac{K}{N} \geq \tilde{R}$

Typical set decoder:

$$\sigma(Y^N) = \begin{cases} \hat{S} & \text{if exactly one } \hat{S} \text{ s.th. } (X^N(\hat{S}), Y^N) \in J_{N,\varepsilon} \\ \perp & \text{otherwise} \end{cases}$$

↑ will choose later

How well does this work? Enough to show that

$$E[P_B] = \frac{1}{2^K} \sum_s \Pr(\hat{S} \neq s \mid S = s) \longrightarrow 0$$

- $E[P_B]$: average over random source message + channel output
- $\Pr(\hat{S} \neq s \mid S = s)$: with respect to channel AND code! — independent of s by symmetry of construction
- average over random choice of code!

Indeed, if true on average for random codes then $\exists$ codes w/ this property!

When is $\hat{S} \neq s$? Recall: $s \longrightarrow X^N(s) \longrightarrow Y^N \longrightarrow \hat{S} = \sigma(Y^N)$.

Two options for errors:

* $(X^N(s), Y^N) \notin J_{N,\varepsilon}$: $\Pr(\cdots) \longrightarrow 0$ by (2)

* $(X^N(s'), Y^N) \in J_{N,\varepsilon}$ for some $s' \neq s$:

$K = \lceil N\tilde{R} \rceil \leq N(\tilde{R} + \frac{1}{N})$

$$\Pr(\ldots) \underset{(3)}{\leq} \#\{s' \neq s\} \cdot 2^{-N(I(X:Y)-3\varepsilon)} \leq 2^{N(\tilde{R}+\frac{1}{N}-I(X:Y)+3\varepsilon)}$$

$\longrightarrow 0$ if we choose ε s.th. $\tilde{R} < I(X:Y) - 3\varepsilon$

$\Longrightarrow \Pr(\hat{S} \neq s \mid S = s) \longrightarrow 0$ for each s, so also $E[P_B] \longrightarrow 0$ (and all at same speed) $\square$

We'll continue here on Monday

Proof of Converse (B) (NOT in Mackay)

"If $\tilde{R} > C_Q(Q)$: $\exists \delta > 0$ $\exists N_0$ $\forall N \geq N_0$: $\exists$ code with $\frac{k}{N} \geq \tilde{R}$ & $p_B \leq \delta$"

Tools: ① Data Processing Inequality (DPI) for $A \to B \to C$ Markov chain:

$$I(A:B) \geq I(A:C) \quad \& \quad H(A|B) \leq H(A|C)$$

Same for $A \to C$ uncorrelated! i.e. $P(a,b,c) = P(a)P(b|a)P(c|b) = P(b)P(a|b)P(c|b)$

② If X^N arbitrary and Y^N channel output: ← i.e. $P(x^N, y^N) = P(x^N)\, Q(y_1|x_1) \cdots Q(y_N|x_N)$

$$I(X^N:Y^N) \leq \sum_{i=1}^{N} I(X_i:Y_i) \leq N \cdot C_Q(Q)$$

HW 5

③ Fano's inequality for $S \to T \to \hat{S}$ Markov chain, $p = P_r(S \neq \hat{S})$

$$H(\{p, 1-p\}) + p \cdot \log \#\mathcal{A}_S \geq H(S|\hat{S}) \geq H(S|T)$$

Ex

Proof of the converse: Consider (N,k)-code with $\frac{k}{N} \geq \tilde{R} > C_Q$.

Let $S \in \{1, \dots, 2^k\}$ uniform. Recall: $S \to X^N \to Y^N \to \hat{S}$.

Then:

$$* \quad H(S|Y^N) = H(S) - I(S:Y^N) \overset{\text{DPI ①}}{\geq} H(S) - I(X^N;Y^N) \overset{②}{\geq} k - N \cdot C_Q$$

($S \to X^N \to Y^N$ Markov chain)

$$* \quad H(S|Y^N) \overset{\text{Fano ③}}{\leq} 1 + P_r(\hat{S} \neq S) \cdot \log \#\mathcal{A}_S = 1 + p_B \cdot k$$

($S \to Y^N \to \hat{S}$ Markov chain)

$$\Rightarrow \quad k - N \cdot C_Q \leq 1 + p_B \cdot k$$

$$\Rightarrow \quad p_B \geq \frac{1}{k}(k - N \cdot C_Q - 1) = 1 - \frac{N \cdot C_Q}{k} - \frac{1}{k} \geq \underbrace{1 - \frac{C_Q}{\tilde{R}}}_{>0} - \underbrace{\frac{1}{N\tilde{R}}}_{\to 0}$$

Can never go below this for large enough N $\square$

Are we happy? What questions does Shannon's theorem leave unaddressed? algorithmics, large N, ... how to even compute C_Q?

Pf of Fano:

Define $E = \begin{cases} 1 & \hat{S} \neq S \\ 0 & S = S \end{cases}$ s.th. $H(E) = H(\{p, 1-p\})$

Use chain rule in two ways:

$$H(ES|\hat{S}) \overset{\text{chain rule}}{=} H(S|\hat{S}) + \underbrace{H(E|S\hat{S})}_{=0} = H(S|\hat{S}) \overset{\text{DPI}}{\geq} H(S|Y)$$

$$H(ES|\hat{S}) \overset{\text{def}}{=} H(E|\hat{S}) + H(S|E\hat{S})$$

$$\leq H(E) + p\, H(S|\hat{S}, E=1) + (1-p) \underbrace{H(S|\hat{S}, E=0)}_{=0 \text{ since } S=\hat{S} \text{ if } E=0}$$

$$\leq H(E) + p \cdot \log \#\text{vals}$$

□