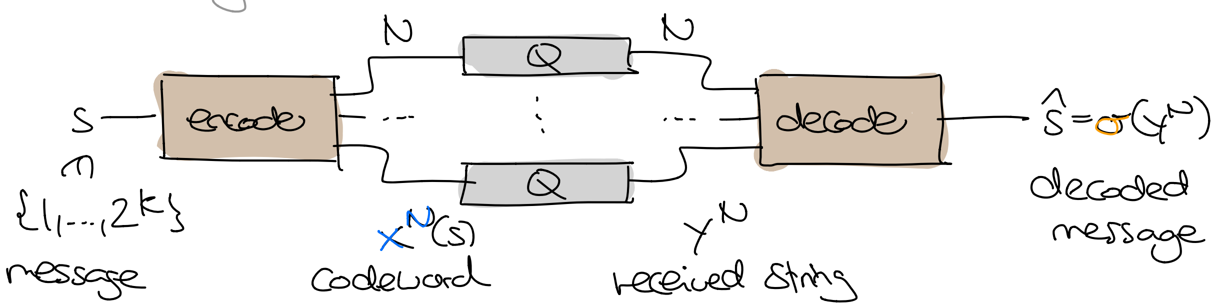
Proof of the Noisy Coding Theorem (§10)

Recall from Tuesday:



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

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

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:

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

enough to prove for P_B

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

Ⓐ Achievability:

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

Ⓑ Converse:

If $\tilde{R} > C(Q)$: $\exists \delta > 0$ $\exists N_0\ \forall N \geq N_0$: $\nexists$ 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 (A)

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

① $\# J_{N,\varepsilon} \leq 2^{N(H(XY)+\varepsilon)}$

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

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

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

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

$\leq 2^{-N(I(X:Y)-3\varepsilon)}$ □

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) ($k = k(N)$) with $\frac{k}{N} \geq \tilde{R}$ s.th. $P_B \xrightarrow{N \to \infty} 0$

Can always upgrade to P_{Bm} 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:

$$\begin{matrix} X^N(1) & = & X_1(1) & X_2(1) & \cdots & X_N(1) \\ \vdots & & \vdots & & & \vdots \\ X^N(2^K) & = & X_1(2^K) & X_2(2^K) & \cdots & X_N(2^K) \end{matrix} \quad \overset{\text{IID}}{\sim} P(x)$$

codeword by codeword, letter by letter

↳ (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

average over random choice of code!

$$E[p_B] = \frac{1}{2^K} \sum_s \underbrace{\Pr(\hat{S} \neq s \mid S = s)}_{\text{independent of } s \text{ by symmetry of construction}} \longrightarrow 0$$

with respect to channel AND code!

average over random source message + channel output

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 $E[p_B] \longrightarrow 0$ $\square$

(and all at same speed)