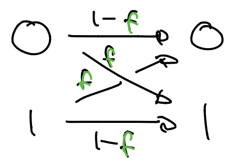
The Noisy Coding Theorem (§9-10)

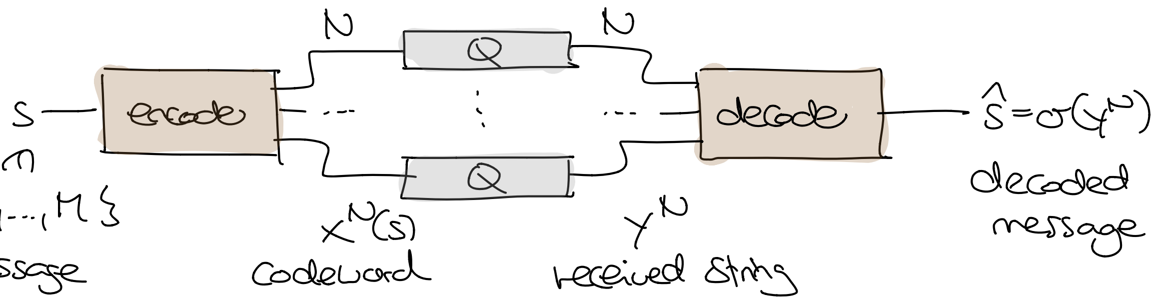
Recall: Capacity of channel $Q(y|x)$:

$$\boxed{C(Q) = \max_{P(x)} I(X:Y)}$$ ⟵ Computed for $P(x,y) = P(x)Q(y|x)$

e.g. $C = 1 - H(\{f, 1-f\})$ for binary symmetric channel



The noisy coding theorem states: The capacity is the "optimal" rate at which we can communicate "reliably" using Q. Let's state this more precisely:



(N,K)-block code: $\quad x^N : \{1, 2, \ldots, M\} \longrightarrow \mathcal{A}_X^N \quad$ where $\quad M = 2^K$

Decoder: $\quad \sigma : \mathcal{A}_Y^N \longrightarrow \{\perp, 1, \ldots, M\}$

⟵ Convenient to indicate failure (but can also just decode incorrectly)

→ distribution of decoded message when sending s:

$$P(\hat{s}|s) = \Pr(\hat{S} = \hat{s} \mid S = s) = \sum_{\substack{y^n \text{ s.th} \\ \sigma(y^n) = \hat{s}}} Q(y_1|x_1(s)) \cdots Q(y_n|x_n(s))$$

Components of $x^N(s)$

Figures of merit:

* rate: $R := \frac{K}{N}$ bits per channel use

* average prob. of (block) error for uniform $S \in \{1, \ldots, M\}$:

$$p_B = \Pr(\hat{S} \neq S) = \frac{1}{M} \sum_{s=1}^{M} \sum_{\hat{s} \neq s} P(\hat{s}|s)$$

Similarly for general $P(s)$

* maximal probability of (block) error:

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

How are these related?

* Clearly: $\boxed{P_{BM} \geq P_B}$

* Conversely: Define $(N, K-1)$-code by removing the $\frac{M}{2} = 2^{K-1}$ codewords with largest $\Pr(\hat{S} \neq S \mid S = s)$. "expurgation"

$$\Longrightarrow \boxed{P_{BM}^{new} \leq 2 P_B \quad \& \quad R^{new} = R - \frac{1}{N}}$$

Pf: Otherwise, original code had $> \frac{M}{2}$ codewords with $\Pr(\hat{S} \neq S \mid S = s) > 2 P_B$

$$\Longrightarrow P_B = \frac{1}{M} \sum_s \Pr(\hat{S} \neq S \mid S = s) > \frac{1}{2} \cdot 2 P_B = P_B \quad \lightning \qquad \square$$

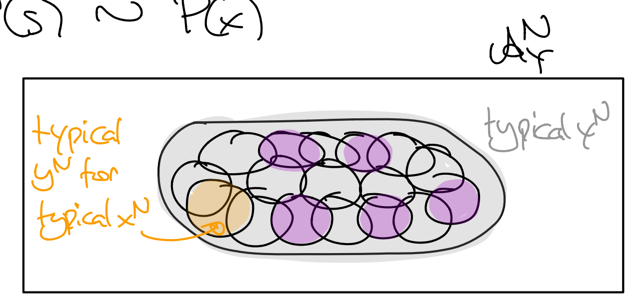
enough to show for for P_B instead of P_{BM}

Shannon's noisy coding theorem: Let $Q(y|x)$ channel and $0 < \delta < 1$.

(A) If $R < C(Q)$: $\exists N_0 \, \forall N \geq N_0$: $\exists (K,N)$-code & decoder with $\frac{K}{N} \geq R$ and $P_{BM} \leq \delta$

(B) ? Thursday!

Intuition: Choose random codewords $X^N(s) \overset{IID}{\sim} P(x)$



typical channel outputs = ?

- in total $\sim 2^{N H(Y)}$
- for typical codeword $\sim 2^{N H(Y|X)}$ not so clear?

↳ can choose $\sim 2^{N H(Y)} / 2^{N H(Y|X)} = 2^{N I(X:Y)}$ also not so clear? with little overlap

↳ do this for $P(x)$ that achieves capacity Cf. noisy typewriter!

Let's make this precise ...

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

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

e.g. $\left| \frac{1}{N} \log \frac{1}{P(x^N, y^N)} - H(X,Y) \right| \leq \varepsilon$

Properties:

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

(by definition) $2^{-N(H(XY)+\varepsilon)} \leq P(x^N, y^N) \leq 2^{-N(H(XY)-\varepsilon)}$

① $\#J_{N,\varepsilon} \leq 2^{N(H(XY)+\varepsilon)}$ (even holds for $T_{N,\varepsilon}(P_{XY})$)

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

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

← — X_i & Y_i correlated via $P(X,Y)$

Pf: $\Pr((X^N, Y^N) \notin J_{N,\varepsilon}) = \Pr(X^N \notin T_{N,\varepsilon}(P_X)$ OR -- OR ...$)$

$\leq \Pr(X^N \notin T_{N,\varepsilon}(P_X)) + \ldots + \ldots$ and each term $\longrightarrow 0$.

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

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

← $\tilde{X}_i$ indep. from $\tilde{Y}_i$

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

$\leq 2^{-N(I(X:Y)-3\varepsilon)}$ □

On Thursday we will use this to prove the noisy coding theorem!