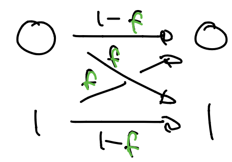
The Noisy Coding Theorem (§9-10)

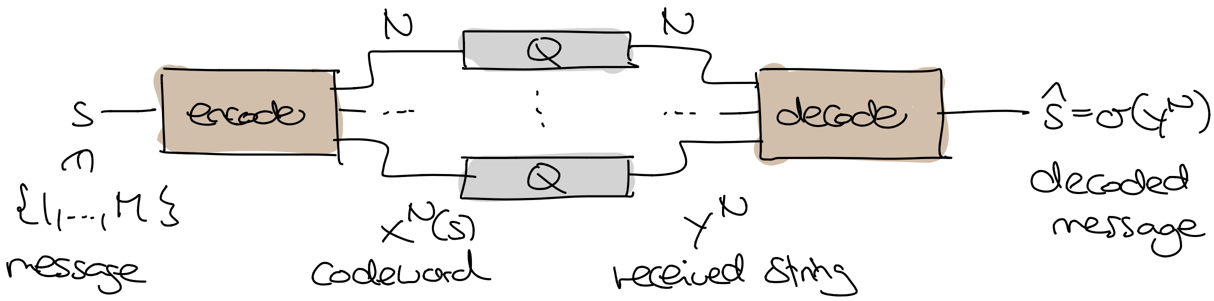
Recall: Capacity of channel $Q(y|x)$:

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

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:



<u>(N,K)-block code:</u> $x^N : \{1, 2, \dots, M\} \longrightarrow \mathcal{A}_X^N$ where $M = 2^K$

<u>Decoder:</u> $\sigma : \mathcal{A}_Y^N \longrightarrow \{\perp, 1, \dots, M\}$

$\perp$: 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 $x^N(s)$

<u>Figures of merit:</u>

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

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

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

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 $\Pr(\hat{S} \neq S \mid S = s) > 2 P_B$

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

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

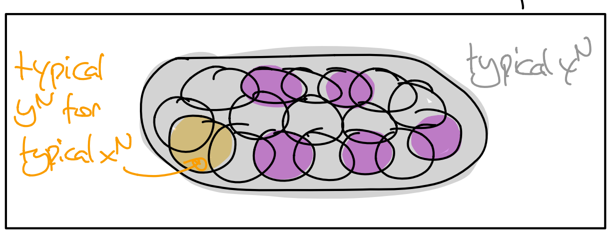
$\boxed{\text{Shannon's noisy coding theorem:}}$ Let $Q(y|x)$ channel and $0 < \delta < 1$.

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

(B) ? Thursday!

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

$\mathcal{A}_Y^N$



\# typical channel outputs = ?

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

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

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

Let's make this precise ...

$\boxed{\text{Jointly typical set}}$ 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 $(x^N, y^N) \in J_{N,\varepsilon}$: $2^{-N(H(\underset{Y}{X})+\varepsilon)} \le P(\underset{y^N}{x^N}) \le 2^{-N(H(\underset{Y}{X})-\varepsilon)}$

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

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

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

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

← — X_i & Y_i correlated via $P(x,y)$

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

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

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

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

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

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

$$\le 2^{-N(I(X:Y)-3\varepsilon)}$$ □

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