Converse of the Noisy Coding Theorem (NOT in MacKay)

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

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

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

Same for $A \leftrightarrow C$ interchanged!

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. $P(x^N, y^N) = P(x^N)\, Q(y_1|x_1) \cdots Q(y_N|x_N)$

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

HW 5

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

$$\boxed{H(\{p, 1-p\}) + p \cdot \log \#A_S \geq H(S|\hat{S}) \underset{①}{\geq} H(S|T)}$$

EX

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

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

Then:

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

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

$* \; H(S|Y^N) \overset{\text{Fano ③}}{\leq} 1 + \Pr(\hat{S} \neq S) \cdot \log \#A_S = 1 + p_B \cdot K$

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

$$\Rightarrow \; K - N \cdot C \leq 1 + p_B \cdot K$$

$$\Rightarrow \; p_B \geq \frac{1}{K}(K - N \cdot C - 1) = 1 - \frac{N \cdot C}{K} - \frac{1}{K} \geq \underbrace{1 - \frac{C}{\tilde{R}}}_{>0} - \underbrace{\frac{1}{N\tilde{R}}}_{\to 0} \qquad \square$$

Can never go below this for large enough N

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

Shannon's Theorem vs. Practice (§11)

- Need large block size N for joint typicality vs. fixed packet size
- Codebook $x^N(1), \dots, x^N(2^k)$ exponentially large in N (if $R>0$) → HW 5
- Random codes vs predictable performance

A family of codes is "very good" if $\frac{k}{N} \to C$ & $P_B \to 0$

"good" if $\frac{k}{N} \geq \tilde{R}$ & $P_B \to 0$ for some $\tilde{R} > 0$

"bad" otherwise

... and "practical" if $\exists$ efficient encoder + decoder

Often run in embedded devices (cell phone, satellite, TV, ...)!

In practice:

do you know what this means?

* most codes are linear (x^N linear function of s^k)

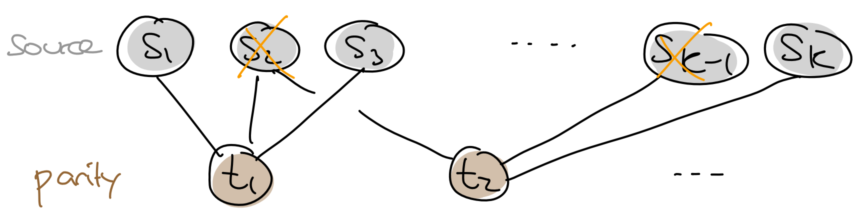
* "easy" to come up with "plausible" encoders — but optimal decoding is in general (NP) hard! ← unlike for compression!

$$\boxed{\sigma_{opt}(y^N) = \arg\max_{\hat{s}} P(\hat{s} \mid y^N)}$$

Why? If $P(s)$ arbitrary prior, want to choose σ to maximize $\Pr(\hat{S} = S)$

$$= \underbrace{\sum_{y^N} \Pr(S = \sigma(y^N), Y^N = y^N)}_{\text{choose } s = \sigma(y^N) \text{ that maximizes } P(s, y^N) \propto P(s \mid y^N)}$$

e.g. imagine the following (LDPC) code:



For erasure channel:

$s_1 \oplus \cancel{s_2} \oplus s_3 \oplus t_1 = 0$

$\cancel{s_2} \oplus \cancel{s_{k-1}} \oplus s_k \oplus t_2 = 0$

...

* types of decoders: "algebraic" vs. "iterative"

Types of codes:

* block codes: e.g. Hamming, Reed-Solomon, LDPC codes

storage, bar codes, sat comm

→ Wednesday

WiFi, DVB, ...

* Convolutional: e.g. turbo codes

3G/4G/LTE, sat. comm.

≈ linear streaming codes