Last time:

Theorem [Shannon's noisy coding theorem]

$$C(N) = \max_{p(x)} I(X:Y)$$

How to prove this?

1. Direct coding – consider high-rate codes

   Not easy -- instead, consider "random" (M,n) codes with rate = I(X:Y) and show Prob($EP_e \to 0$) > 0.

   Thus $\exists$ code with small EP_e (our 2nd encounter with "existential proofs"). Extract a subcode with similar rate but $P_e \to 0$.

2. Converse – show that at higher rates, $EP_e \not\to 0$.

Plan: 2, then heuristic 1, then 1.

1. Direct coding:

$$
\begin{aligned}
c_1 &= x_{11}, x_{12}, \ldots, x_{1n} \\
c_2 &= x_{21}, x_{22}, \ldots, x_{2n} \\
&\;\;\vdots \\
c_M &= x_{M1}, x_{M2}, \ldots, x_{Mn}
\end{aligned}
$$

x_{ij} chosen iid $\sim p(x)$

Better proof why $P_e \to 0$.

Long version (18 pages) available in homepage, here, a 6-page version skipping ε, δ & some details.

Recall:

Def[typical sequence]:

x^n ε-typical if $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$

It means $2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$.

\begin{General technical tool}

Def[Jointly typical sequence]:

$x^n y^n$ ε-jointly-typical if

$$|-1/n \log(p(x^n y^n)) - H(XY)| \leq \varepsilon$$

where $p(x^n\, y^n) = \Pi_{i=1}^{n}\, p(x_i\, y_i)$.

Need also: (a) $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$
(b) $|-1/n \log(p(y^n)) - H(Y)| \leq \varepsilon$

[The strong typicality has (c) $\Rightarrow$ (a,b), but not for entropic typicality.]

Def[Jointly-typical set]: $A_{n,\varepsilon} = \{x^n y^n$ ε-jointly typical$\}$

Joint asymptotic equipartition (Joint AEP) theorem:

Let (X^n, Y^n) be sequences of length n

drawn iid according to $p(x^n\ y^n) = \Pi_{i=1}^{n}\ p(x_i\ y_i)$.

Then:

1. $Pr(X^nY^n \in A_{n,\varepsilon}) \to 1$
2. $|A_{n,\varepsilon}| \approx 2^{nH(XY)}$
3. if we draw X^n & Y^n according to $q(x^n\ y^n) = p(x^n)\ p(y^n)$.
   $Pr_q\ (\text{outcome} \in A_{n,\varepsilon}) \approx 2^{-nI(X:Y)}$

Proved in the 18 page notes, similar to the proof for the asymptotic equipartition thm.

More observations:

Given $y^n \in T^Y_{n,\varepsilon}$, collect in a set $S(y^n)$
all those $x^n \in T^X_{n,\varepsilon}$ s.t. $x^n y^n \in A_{n,\varepsilon}$.

(1) $p(x^n|y^n) = p(x^n y^n) / p(y^n) \approx 2^{-n[H(XY)-H(Y)]} = 2^{-n[H(X|Y)]}$

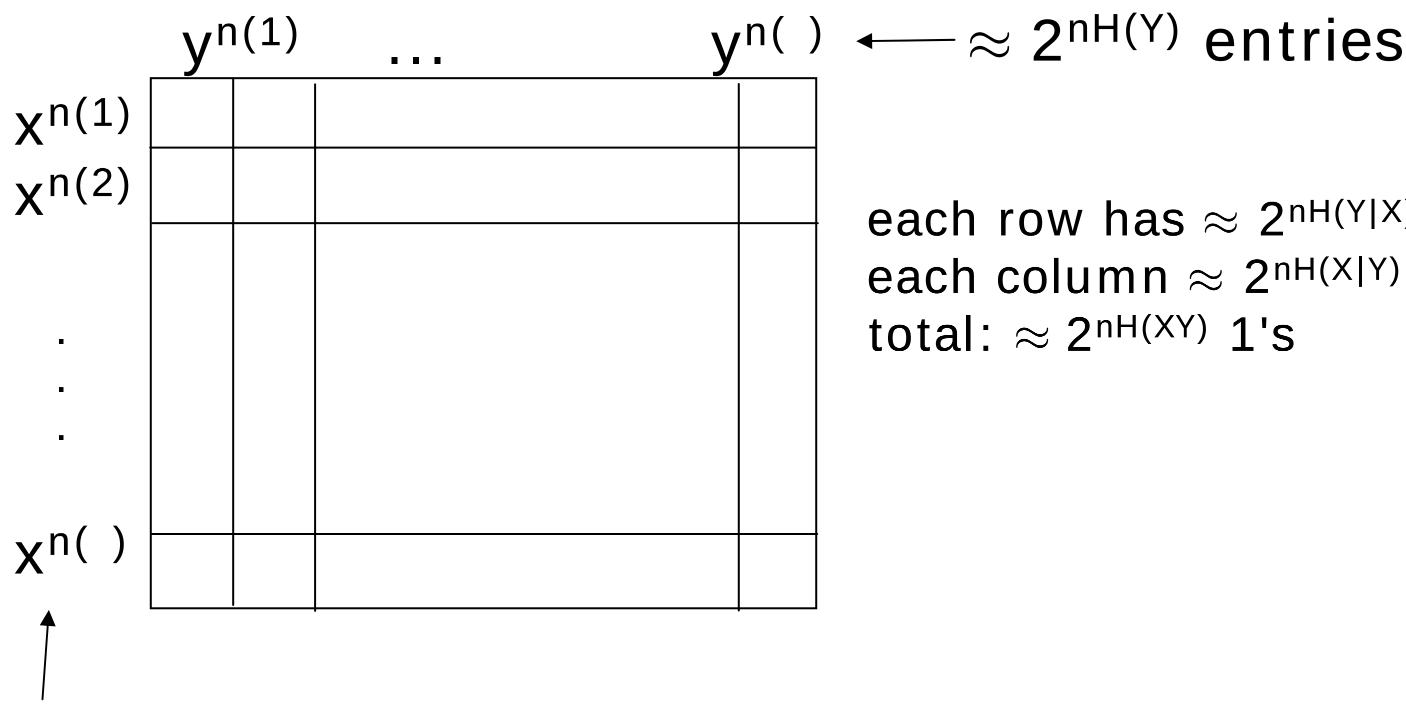
$\uparrow$ since $x^n y^n \in A_{n,\varepsilon}$!!

(2) $1 = \sum_{x^n \in S} p(x^n|y^n) \approx |S(y^n)|\ 2^{-n[H(X|Y)]}$

Hence, $|S(y^n)| \approx 2^{nH(X|Y)}$. Fraction of such $x^n \approx 2^{-nI(X:Y)}$.

Similarly, given $x^n \in T^X_{n,\varepsilon}$, $\approx 2^{nH(Y|X)}$ y^n's are jointly typical with it, and the fraction of such $y^n \approx 2^{-nI(X:Y)}$.

Make a table of typical x^n's and y^n's, and for jointly typical $x^n y^n$, put a 1, else, put a 0.



each row has $\approx 2^{nH(Y|X)}$ 1's
each column $\approx 2^{nH(X|Y)}$ 1's
total: $\approx 2^{nH(XY)}$ 1's

$\approx 2^{nH(X)}$ entries

\end{General technical tool}

Our random code corresponds to M randomly chosen rows.

Back to the direct coding proof:

D_n: typical set decoding

Given y^n:

If there is a unique $x^n \in S(y^n)$ output m' s.t. $c_{m'} = x^n$.

Else, output W=M+1 (error symbol).

How will this fail for input message m?

Either - no such m' $\quad \mathrm{Err}_0$ unlikely–only when $x^n y^n$ not jointly typical

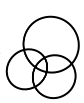
- or $\exists m'' \neq m$ with $c_{m''} y^n \in A_{\varepsilon,n}$ $\quad \mathrm{Err}_{m''}$

For the random code $\mathcal{C}_n$, let $EP_e(\mathcal{C}_n)$ be the average error (over all messages), by symmetry, same as error for m=1.

Averaging over the choice of $\mathcal{C}_n$:

$$\Pr_{\mathcal{C}_n} EP_e(\mathcal{C}_n) = \Pr_{\mathcal{C}_n}(W \neq 1 \mid m=1) = \Pr_{\mathcal{C}_n}(\underset{\uparrow \text{ unlikely}}{\mathrm{Err}_0} \cup \underbrace{\mathrm{Err}_2 \ldots \cup \mathrm{Err}_M}_{\text{equiprobable}} \mid m=1)$$

union bdd



$$\leq M \Pr_{\mathcal{C}_n}(\mathrm{Err}_2 \mid m=1)$$

Bounding $\Pr_{\mathcal{C}_n}(\mathrm{Err}_2 | m=1) = \Pr_{\mathcal{C}_n}(c_2 y^n \in A_{n,\varepsilon_n})$:

But c_2 and $y^n = N^{\otimes n}(x_1)$ independent.

By joint AEP (3), $\Pr_{\mathcal{C}_n}(c_2 y^n \in A_{n,\varepsilon_n}) \approx 2^{-nI(X:Y)}$

If $M = 2^{n(I(X:Y) - \delta_n))}$ and $n\delta_n$ grows with n but $\delta_n \to 0$

$$\Pr_{\mathcal{C}_n} EP_e(\mathcal{C}_n) \leq M \Pr_{\mathcal{C}_n}(\mathrm{Err}_2 | m=1) \to 0$$

Some code $\mathcal{C}_n$ (in fact most codes) has vanishing $EP_e(\mathcal{C}_n)$.

Fix a code $\mathcal{C}_n$ that has vanishing $EP_e(\mathcal{C}_n)$.

Claim:
Expunging the worse half of the codewords from $\mathcal{C}_n$,
we get a new code $\mathcal{C}'_n$ with $P_e(\mathcal{C}'_n) \leq 2\, EP_e(\mathcal{C}_n)$.

Proof:

Reorder m's so that $P_e(m)$ is increasing.

$$\underbrace{P_e(1) + P_e(2) + \ldots + P_e(M/2)}_{\text{replace each by zero}} + \underbrace{P_e(M/2+1) + \ldots + P_e(M)}_{\text{replace each by } P_e(M/2)} = M\, EP_e(\mathcal{C}_n)$$

So, $M/2\; P_e(M/2) \leq M\, EP_e(\mathcal{C}_n)$, $P_e(M/2) \leq 2\, EP_e(\mathcal{C}_n)$.

Keeping only codewords for $m=1, \ldots, M/2$,
worse case prob error $= P_e(M/2) \leq 2\, EP_e(\mathcal{C}_n)$.
Rate decreases only by $1/n$.