Lecture 2

repetition code reduces order of error from p to p^2.

we can then concatenate the code to reduce order even more, i.e.

$0 \to 000 \to 000\ 000\ 000$

$1 \to 111 \to 111\ 111\ 111$

rate goes to $\frac{1}{3^2} = \frac{1}{9}$

& error prob. is

$$3\bar{p}^2 - 2\bar{p}^3$$

where $\bar{p} = 3p^2 - 2p^3$

$\Rightarrow$ order of error is p^4

In general, after n concatenations, rate goes to $\frac{1}{3^n}$ & error order is $O(p^{2n})$.

Thus, we can make error probability arbitrarily small at the cost of the rate converging exponentially fast to zero.

Is there a better way?

Yes, this was a major achievement of Shannon.

- Suppose that a noisy classical channel is described by a conditional probability distribution $P_{Y|X}(y|x)$.

(generalizes binary symmetric channel)

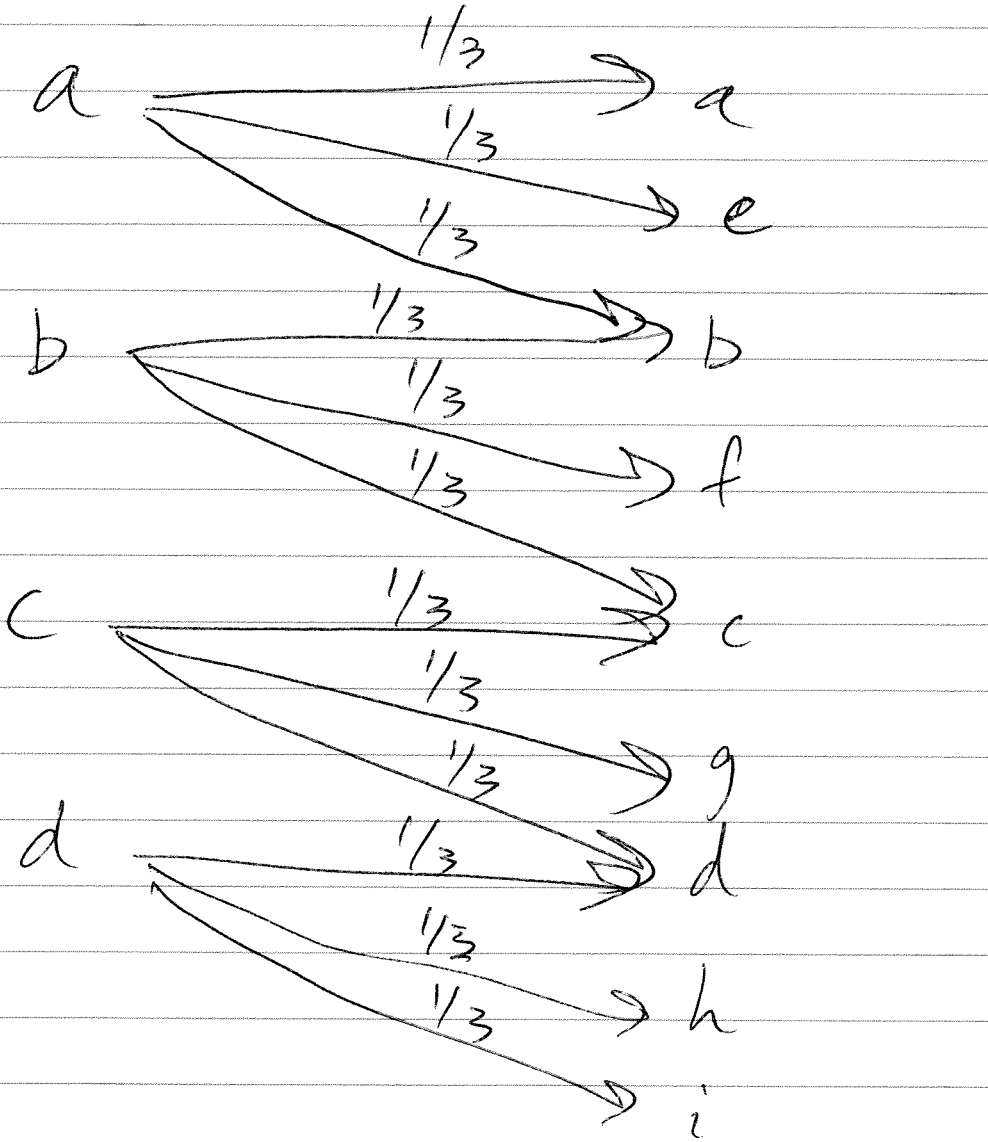
- Then we want to know the maximum # of bits per channel use we can transmit, such that error prob. tends to zero as # of channel uses becomes large.

- This quantity is called channel capacity.

Assumption - channel is called in i.i.d. way

Another simple example to motivate idea behind channel capacity theorem:

"noisy typewriter" channel



How to code for this channel to recover input perfectly?

Use only a subset of inputs, i.e., $\{a, g, c\}$

Then we can decode error free.

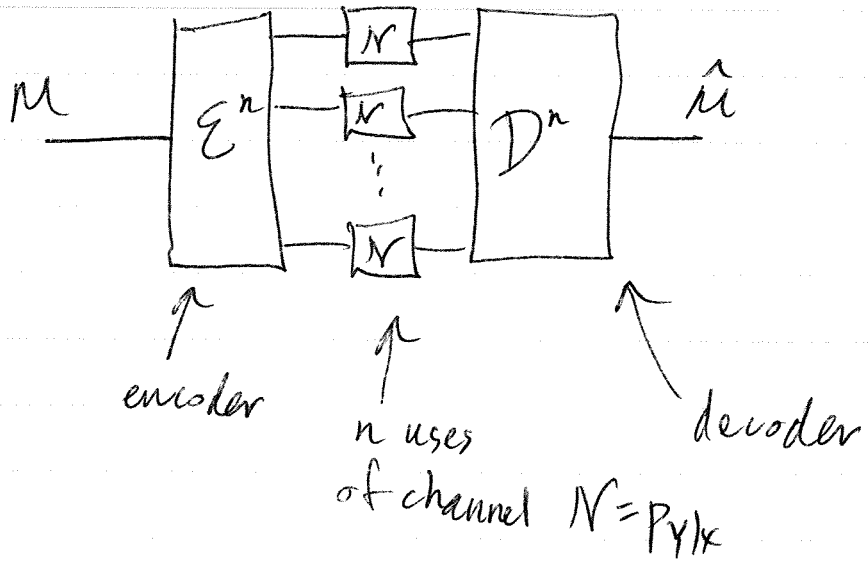
If receiver gets a, e, b, then decode as a.

If receiver gets c, g, d, then decode as c.

Impossible to receive f, c, h, i

generic i.i.d. situation not too different from this.

General setting of communication



goal: $\Pr[\hat{M} \neq M] \leq \varepsilon$

$\forall$ possible distributions on M

rate $= \dfrac{\log_2 |M|}{n} = \dfrac{\text{bits}}{\text{channel use}}$

Two ideas for achieving channel capacity:

1) conditional typicality

2) random coding

Given a pair (x^n, y^n) of sequences, define y^n to be δ-conditionally typical if

$$|\bar{H}(y^n|x^n) - H(Y|X)| \leq \delta$$

where $\bar{H}(y^n|x^n)$ is the conditional sample entropy, defined as

$$\bar{H}(y^n|x^n) \equiv -\frac{\log_2 p_{Y^n|X^n}(y^n|x^n)}{n}$$

& distributions $p_X(x)$ & $p_{Y|X}(y|x)$

& the true conditional entropy is defined as

$$H(Y|X) = \sum_x P_X(x) \left(-\sum_y P_{Y|X}(y|x) \log_2 P_{Y|X}(y|x) \right)$$

$$= \sum_x P_X(x) \; H(Y|X=x)$$

can prove that $H(Y|X) = H(XY) - H(X)$

Conditionally typical set is then

$$T_\delta^{Y^n|x^n} = \{ y^n \in \mathcal{Y}^n : |\bar{H}(y^n|x^n) - H(Y|X)| \leq \delta \}$$

From LLN, we have

$$\mathbb{E}_{X^n} \left\{ \Pr_{Y^n|X^n} \left\{ Y^n \in T_\delta^{Y^n|X^n} \right\} \right\} \geq 1 - \varepsilon$$

$\forall \varepsilon \in (0,1)$, $\delta > 0$ & suff. large n.

Also $\left| T_\delta^{Y^n|x^n} \right| \leq 2^{n[H(Y|X)+\delta]}$

Recall that we defined mutual information of RVs X & Y as

$$I(X;Y) = H(X) + H(Y) - H(XY)$$
$$= H(Y) - H(Y|X)$$

we argued that $I(X;Y) \geq 0$, which implies that

$$H(Y) \geq H(Y|X)$$

Random coding: Shannon's idea:

Analyzing the error probability for the best possible code seems difficult.

Instead pick the code @ random & analyze error probability in this way. If error probability can be made small, then deduce that there exists a code achieving capacity.

To pick a code @ random,
use distribution $p_X(x)$.

Message	symbols			
1	$x_1(1)$	$x_2(1)$	$\cdots$	$x_n(1) \equiv x^n(1)$
2	$x_1(2)$	$x_2(2)$	$\cdots$	$x_n(2) = x^n(2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
M	$x_1(M)$	$x_2(M)$	$\cdots$	$x_n(M) = x^n(M)$

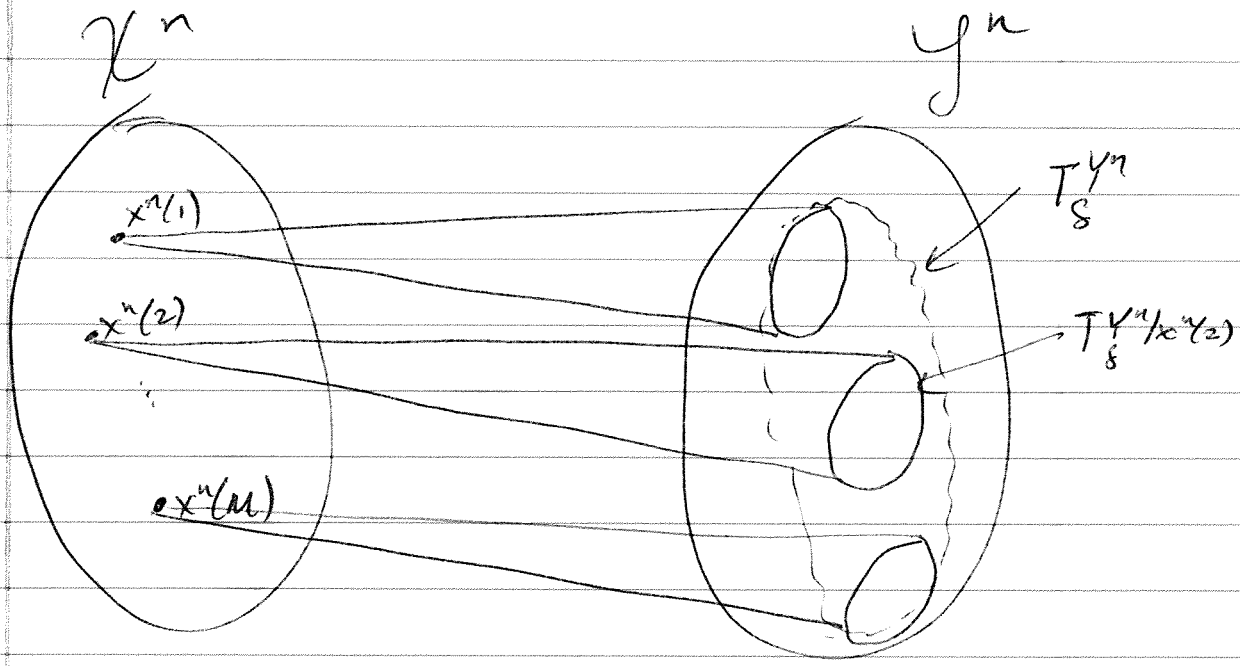
every symbol chosen randomly
according to $p_X(x)$

comment: might seem like a stupid
way of picking a code.

Indeed, when n is small,
this won't work well.

However, this will work well for
large n, as long as message set
is not too large.

the basic intuition for the coding scheme is given by this picture



for a random choice of x^n, if we do not know the value, then we average over it, & the distribution is the i.i.d. version of

$$P_Y(y) = \sum_x P_X(x)\, P_{Y|X}(y|x)$$

It is then highly likely that y^n is in the typical set $T_\delta^{Y^n}$ w/ size $\approx 2^{nH(Y)}$

If the input x^n is known,
then it is highly likely
that $y^n \in T_\delta^{Y^n|x^n}$ (i.e., that it lands in the conditionally typical set)

The question then is:

How many codewords can we have such that the outputs are distinguishable? (such that the output conditionally typical sets are non-overlapping)

this is related to a sphere packing problem, & the answer is intuitively

$$\frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{nI(X;Y)}$$

So the # of messages that can be faithfully transmitted is $|M| = 2^{nI(X;Y)}$ w/ rate equal to

$$\frac{\log_2 |M|}{n} = I(X;Y)$$

the scheme involved picking codewords from $P_X(x)$.

We can optimize the rate by calculating $\max_{P_X(x)} I(X;Y)$

& this can be shown to be the capacity of the channel $P_{Y|X}$