Lec 5 May 20, 2010

Note Title 19/05/2010

Def: A (classical) channel N is specified by:

- input alphabet X
- output ..... Y
- a distribution $p(y|x)$ for each $x \in X$.

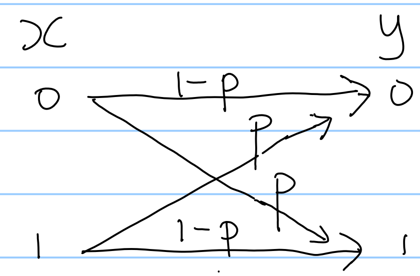
$$\xrightarrow{x \in X} \boxed{N} \xrightarrow{y \in Y} \quad \text{wp } p(y|x)$$

Aside: Can write N as a stochastic matrix

$$\begin{array}{cc} & \longleftarrow \text{all possible } x \longrightarrow \\ \text{all possible } y \ \updownarrow & \begin{bmatrix} & p(y_1|x) & & \\ & p(y_2|x) & & \vdots \\ & \vdots & & \end{bmatrix} \end{array}$$

eg 1 Binary symmetric Channel (BSC)

$X = Y = \{0,1\}$, input $\begin{cases} \text{sent wp } 1-p \\ \text{flipped wp } p \end{cases}$



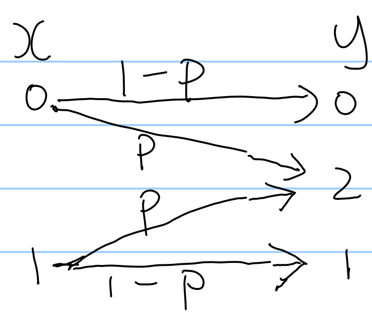
$$\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-p & p \\ 1 & p & 1-p \end{array}$$

eg 2 Erasure channel (E_p)

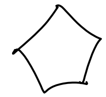
$X = \{0,1\}$,

$Y = \{0,1,2\}$,

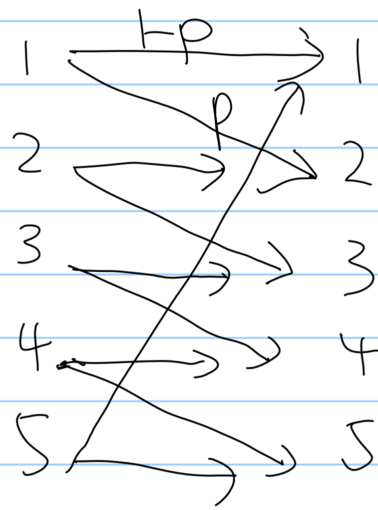
input $\begin{cases} \text{sent wp } 1-p \\ \text{replaced by 2 wp } p \end{cases}$



$$\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-p & 0 \\ 1 & 0 & 1-p \\ 2 & p & p \end{array}$$

eg 3. Pentagon channel 

$X = Y = \{1, 2, 3, 4, 5\}$, input $\begin{cases} \text{sent w.p. } 1-p \\ \text{shifted up mod } 5 \text{ w.p. } p \end{cases}$



$$\begin{array}{c|c} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 1-p & 0 & \cdots & & p \\ p & 1-p & \ddots & & \\ 0 & p & & \ddots & \\ 0 & 0 & & & \ddots \\ 0 & 0 & & & 1-p \end{bmatrix} \end{array}$$

General assumptions:

1. Can use channel many (n) times
2. Each use identical & independent:

   for inputs $x_1 x_2 \dots x_n$

   outputs $= y_1 y_2 \dots y_n$ wp $\prod_{i=1}^{n} p(y_i | x_i)$

"Called discrete memoryless channels DMCs"

Non DMCs:

eg 1 Time vary channel: the ith use is a BSC with prob error p_i

eg 2 Burst error: $X_1 X_2 \dots X_n \longrightarrow X_1 X_2$ ~~////~~ X_n

missing a contiguous block in the output

"Dog eats a page from your book"

eg3 $x_1 x_2 \cdots x_i x_j \cdots x_n$

$\downarrow$

$x_1 x_2 \cdots x_j x_i \cdots x_n$

symbols emerging in slightly wrong order

eg4 $x_1 x_2 \cdots x_n$

$\downarrow$

$y_1 y_2 \cdots y_m$ $\quad m < n$

"Missing messages" – don't know which ones.

Aside: quantum analogues and coding strategies?

DMC from now on

Dealing with noise by error correcting codes:

eg 1. repeat

$$\begin{array}{l} 0 \to \overbrace{00\cdots0}^{\text{k times}} \\ 1 \to 11\cdots1 \end{array} \Bigg\} \text{ majority decoding}$$

messages

a codeword for each message

"The code" = set of code words

= subset of all possible inputs

eg 2. Hamming codes (eg encode 4 bits in 7
corrects up to 1 error)

Each codeword x satisfies 3 parity constraints:
$$x = x_1 x_2 \dots x_7$$

$$P = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad Px = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$ ie $x_1 \oplus x_3 \oplus x_5 \oplus x_7 = 0$

$x_2 \oplus x_3 \oplus x_6 \oplus x_7 = 0$

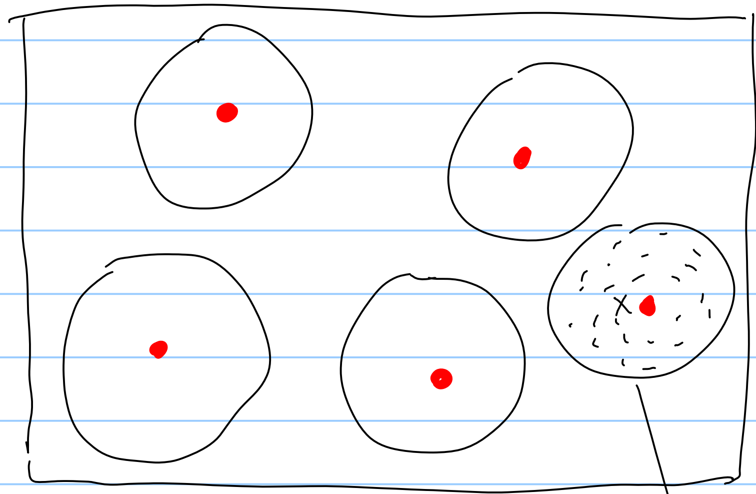
etc

What's cool: if $y_i = x_i + e_i$ and only $e_i = 1$

then $Py = Pe$ = ith col of P,

decoding / identifying the error is easy!

Geometrically: (say $X = Y$)



• Codewords

every output strings up to k errors from x

$$X^{\otimes n} = Y^{\otimes n} \leftarrow n \text{ copies}$$

Can recover message if codewords are sparse enough so that these spheres don't overlap.

Qn: For a fixed message size, to have smaller & smaller
error prob, need bigger & bigger codes ..

(1) that brings more and more errors too

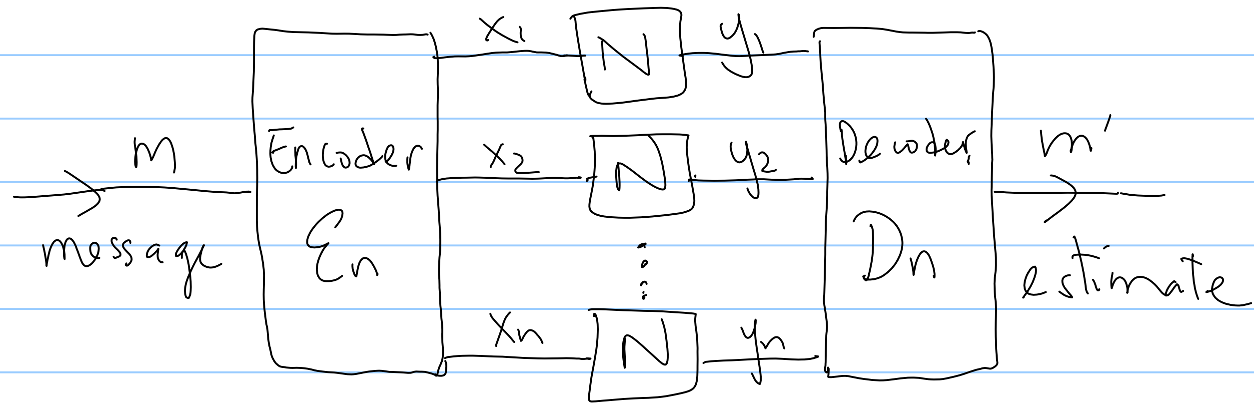
(2) will the rate -> 0 ?

(3) for growing message size,
will prob(every part correct) -> 0?

Usually (1) not a problem if prob error small enough to start with, but (2), (3) can happen, say, with the repetition code.

Will see, we can do much much better
-- magic: iid channel use + large n

Sending messages through n uses of a noisy channel:



An (M,n) code consists of
(1) index set $\mathcal{M} = \{1, \ldots, M\}$
(2) an encoding function $\mathcal{E}_n : \mathcal{M} \longrightarrow X^{\otimes n}$
(3) a decoding function $\mathcal{D}_n : Y^{\otimes n} \longrightarrow \mathcal{M}$

The codewords are $\mathcal{E}_n(1), \mathcal{E}_n(2), \ldots \mathcal{E}_n(M) \leftarrow$ The code

For message m, there's an error if

$$m' = D_n \circ N^{\otimes n} \circ \mathcal{E}_n(m) \neq m$$

Say, happens wp $P_e(m)$

Define P_e^n = worse case prob of error $= \max_{m \in \mathcal{M}} P_e(m)$

$\mathbb{E} P_e^n$ = average . . . . $= \frac{1}{M} \sum_{m=1}^{M} P_e(m)$

Rate of an (M, n) code : $\frac{1}{n} \log M$

Def: For a channel $\mathcal{N}$, a rate R is achievable if $\exists$ sequence of $(M = \lceil 2^{nR} \rceil, n)$ codes s.t. $P_e^n \to 0$ as $n \to \infty$

Def: Capacity of $\mathcal{N}$, $C(\mathcal{N}) = \sup$ over achievable rates

NB If $C > 0$, the entire message, longer & longer ($\propto n$) comes out correctly almost surely!

Thm (Shannon's noisy coding theorem)

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

NB1. $p(xy) = \underset{\substack{\uparrow \\ \text{max over}}}{p(x)}\ \underset{\substack{\uparrow \\ \text{specified by } N}}{p(y|x)}$

NB2. Expression involves only 1 copy of $p(xy)$
but $C(N)$ has an asymptotic definition.

NB3. Works in worse case, no distribution of message
"$p(x)$" in the max has meaning TBD.

NB4. Every channel (but one) has $C > 0$!

eg1. BSC

$$I(X:Y) = H(Y) - \underbrace{H(Y|X)}_{H(p) \text{ indep of } p(x)}$$

max this by making y random
possible when $p(0) = p(1) = \frac{1}{2}$.

$\therefore$ Capacity of BSC $= 1 - H(p)$

eg2 Erasure Channel

$$I(X:Y) = H(X) - \underbrace{H(Y|X)}_{pH(X)} = (1-p)\,H(X)$$

Again optimal
$p(x) : p(0) = p(1) = \frac{1}{2}$.

Same rate as if where the erasures are are known up-front!

Capacity of erasure channel $= (1-p)$

eg 3. Pentagon channel with $p = \frac{1}{2}$

$$I(X:Y) = H(Y) - \underbrace{H(Y|X)}_{\text{always} = 1}$$

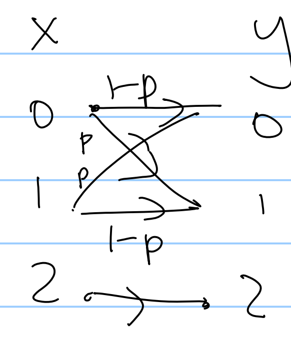
Again optimal $p(x)$ uniform,

$$C(⬠) = \log 5 - 1 = \log\left(\frac{5}{2}\right)$$

eg 1–3 very symmetric

thus optimal $p(x)$ uniform

Try eq 4



? HW

NB If we demand $p_e = 0$, but allowing many uses, we're studing the "zero-error-capacity" (lower bdd for $C(N)$)

eg. The BSC & erasure channel has 0 zero-error capacity

That of ⬠ is $\log\sqrt{5}$, That of eq 4 is 1.

Comparing ⬠ with $E^5_{10^{-10}}$ (erasure channel with $|X| = 5$, $p = 10^{-10}$)

$C(E^5_{10^{-10}}) \approx \log 5 > C(⬠)$ But zero error capacity of $E^5_{10^{-10}} = 0$ $<$ that of ⬠

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