QIC 890/CO781 Lec 09. Oct 06, 2016

Def A classical channel N is specified by:

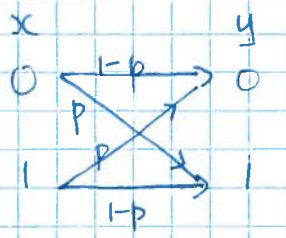
- an input alphabet Ω_X
- an output alphabet Ω_Y
- a distribution $p(y|x)$ for each $x \in X$.

eg 1 Binary symmetric channel (BSC)

$\Omega_X = \Omega_Y = \{0,1\}$

Input is "flipped" wp p, $p(y=x|x) = 1-p$

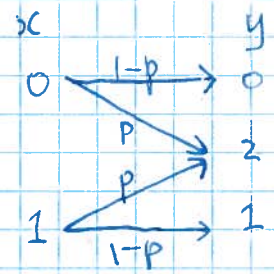
$p(y \neq x|x) = p$



eg 2 Erasure channel

$\Omega_X = \{0,1\}$, $\Omega_Y = \{0,1,2\}$

Input is "erased" (replaced by 2) wp p.



Focus on:

① Asymptotic rate of communication

+ can use channel many (n) times
− allow a very small error prob

② Discrete memoryless channels (DMCs)

ie each use is independent and identical.

When the input is $x_1 x_2 \dots x_n$
the output is $y_1 y_2 \dots y_n$ wp $\prod_{i=1}^{n} p(y_i|x_i)$

- Sensible channels that are out of scope:

eg. Missing-symbol-channel

$x_1 x_2 \dots x_n \to y_1 y_2 \dots y_m$ where $m < n$.

$n-m$ symbols are deleted but we don't know which.

eg. $x_1 x_2 \dots x_n \to x_1 x_2 \dots x_{i-1} x_{i+1} x_i \dots x_n$

Symbols emerge out of order.

eg. Burst errors

$x_1 x_2 \dots x_n \to x_1 x_2$ x_n

Missing a large contiguous block of symbols
like a page is pulled off a book.

- Many recent work on short block length, both classically & quantumly

Assuming DMC, the idea of a "block code" of length n is to restrict the input $x_1 x_2 \cdots x_n$ to a code $C \subseteq \Omega_X^n$.

eg. Repetition code

Let $k \cdot r = n$.

Send r bits, each repeated k times.

$0 \to 0 \cdots 0$ (k times)

$1 \to 1 \cdots 1$ (k times)

Majority decoding.

Code $= \{ b_1^k b_2^k \cdots b_r^k : b_i \in \{0,1\} \}$

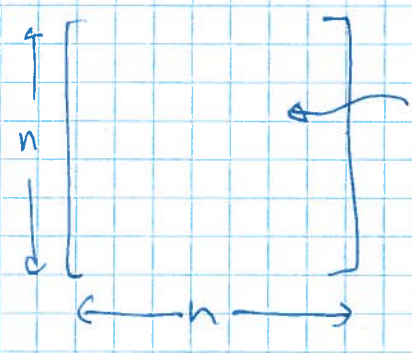
$$\text{Rate} : \frac{\log(\#\text{messages communicated})}{\#\text{channel use} \times \log|\Omega_X|} = \frac{r}{n}$$

Prob of error $\approx p^k$ for erasure channel, for each block

$1 - (1-p^k)^r$ for the entire message.

If $r = \gamma n$, $k = \frac{1}{\gamma}$, error prob $\approx 1 - (1-p^{\frac{1}{\gamma}})^{\gamma n} \to 1$ as $n \to \infty$.

eg. Hadamard code. Start with a Hadamard matrix of order n



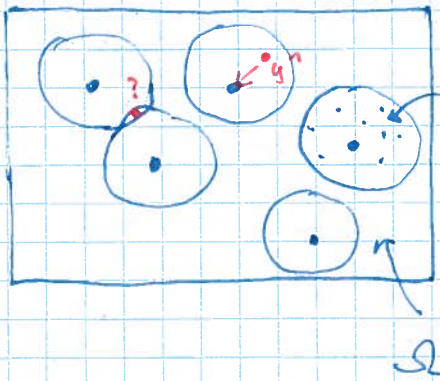
entries ± 1, mutually orthogonal rows

so diff in $\frac{n}{2}$ positions.

can correct up to $\left(\frac{n}{4} - 1\right)$ flips in unknown positions.

Rate $\dfrac{k = \log n}{n} \to 0$ as $n \to \infty$.

Geometric interpretation: ($\mathcal{X} = \mathcal{Y} = \mathcal{Z}$ for simplicity)



• : $x^n \in C$ codeword

Hamming sphere of radius t
= possible strings received by Bob
if no more than t errors occur.

If no more than t errors are likely, desirable to have non-overlapping Hamming spheres of radius t.

Then given y^n in one of these spheres, Bob can learn x^n giving rise to y^n.

i. Limit # codewords & the rates.

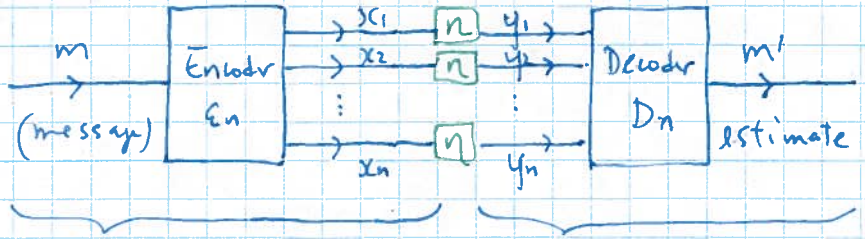
Qn:

To have smaller & smaller errors, need larger & larger block length, which brings more & more errors.

Can prob error $\to 0$ and rate > 0 ?

NB We say there is an error if any part of the message is incorrectly received.

Sending messages through n uses of a noisy channel:



An "(M, n)" code C_n consists of:

① an index set $\mathcal{M} = \{1, 2, \ldots, M\}$

② an encoding function $\mathcal{E}_n : \mathcal{M} \to \Omega_X^n$

③ a decoding function $D_n : \Omega_Y^n \to \mathcal{M}$

} will see why randomness is not needed

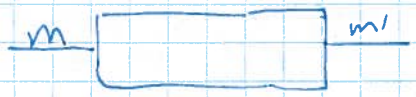
"The code" $= \{\mathcal{E}_n(1), \mathcal{E}_n(2), \ldots, \mathcal{E}_n(M)\} = \{c_1, c_2, \ldots, c_M\}$

codewords

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

For a given (M, n) code C_n:

* for each m, m' is a rv with dist^n $p(m'|m)$



is a channel

* Define $p_e(m) = \text{prob}(m' \neq m \,|\, m) = \text{prob}(D_n \circ \eta^{\otimes n} \circ \mathcal{E}_n(m) \neq m')$ — over channel noise

$p_e(C_n) = \max_{m \in \mathcal{M}} p_e(m)$ worse case error over m

$\bar{p}_e(C_n) = \frac{1}{M} \sum_{m \in \mathcal{M}} p_e(m)$ average case error over m

Def: [achievable rate]

For a channel N, a rate R is achievable if $\exists$ sequence of $(\lfloor 2^{nR} \rfloor, n)$ codes C_n s.t. $P_e(C_n) \to 0$ as $n \to \infty$.

Def: The capacity of N, $C(N)$, is the supremum over achievable rates. } optimal # bits transmitted per use of N

NB: If $C(N) > 0$, the message, which is longer & longer ($\simeq nR$ bits) comes out correctly in each symbol almost surely!

Thm: Shannon's noisy coding theorem

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

where $p(xy) = \underbrace{p(x)}_{\text{optimized over}} \cdot \underbrace{p(y|x)}_{\text{Given by the channel}}$

NB: $C(N)$ is an asymptotic operational quantity

RHS is a "single-letter-formula"; an optimization involving one use of the channel.

NB $p(x)$ is NOT the distribution of the input

We'll see how $p(x)$ comes in.

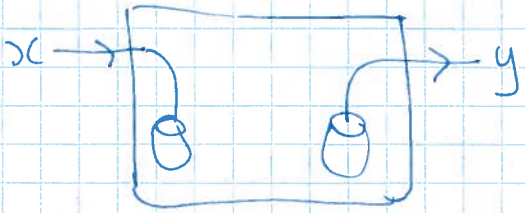
Note the codes work in the worse case.

eg 1 $C(N)=0 \iff \forall x,y,\ p(x,y)=p(x)\cdot p(y)$ (X,Y indep)

$\iff \forall x,y,\ p(y|x)=p(y)$

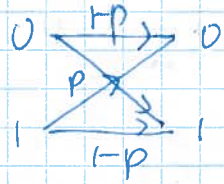
So channel discards input & draws y according to $p(y)$.

We call this channel garbage channel.



All other channels have $C(N)>0$!

eg 2 BSC N_p.



$I(X:Y) = H(Y) - H(Y|X)$

$= \sum_x p(x)\, H(Y|X=x)$

$= \sum_x p(x)\, h(p) = h(p)$ indep of $p(x)$.

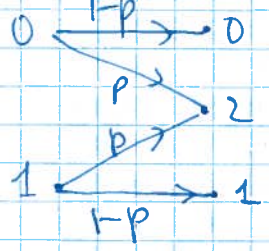
choose $p(x)$ to max $H(Y)$

optimal: $p(Y=0)=p(Y=1)=\frac{1}{2}$ so $p(y=0)=p(y=1)=\frac{1}{2}$

So $H(Y)=1$ (max).

$\therefore C(N_p) = 1-h(p)$

eg 3 Erasure channel $\mathcal{E}_p$.



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

$$p(y=0)\underbrace{H(X|y=0)}_{0} + p(y=1)\underbrace{H(X|y=1)}_{0} + \underbrace{p(y=2)}_{p}\underbrace{H(X|y=2)}_{1}$$

indep of $p(x)$

↳ max this term

with $p(X=0) = p(X=1) = \frac{1}{2}$

$C(\mathcal{E}_p) = 1-p$.

General approach to prove capacity theorems:

(1) Prove a <u>direct coding theorem</u>:

In this case, given any $p(x)$, show

$\exists$ codes achieving the rate $I(X;Y)$

such that $P_e(C_n) \to 0$.

This gives $C(N) \geq \max_{p(x)} I(X;Y)$.

(2) Prove a <u>converse</u>:

For any achievable R, $R \leq \max_{p(x)} I(X;Y)$,

which gives $C(N) \leq \max_{p(x)} I(X;Y)$.

(1) & (2) Generally holds for very different reasons.

Direct coding theorem: fix $p(x)$.

* Need to show $\exists$ (M, n) codes C_n s.t.
  - rate $\frac{1}{n} \log M \geq I(X:Y) - \delta_n$, $\delta_n \to 0$
  - error $P_e(C_n) \to 0$

* Shannon: no need to find these codes.
  Instead, $\forall n$, generate C_n by a random process.
  Show: $\mathop{\mathbb{E}}_{C_n} \bar{P}_e(C_n) \to 0$
  ↑ error averaged over all messages

Then: $\exists \tilde{C}_n$ s.t. $\bar{P}_e(\tilde{C}_n) \to 0$

Then: $\exists \tilde{\tilde{C}}_n$ s.t. $P_e(\tilde{\tilde{C}}_n) \to 0$.

To bound $\mathbb{E}_{C_n} \overline{P_e}(C_n)$:

① Given any $n, M, p(x)$, we generate C_n as follows:

For $i = 1, \ldots, M$
$j = 1, \ldots, n$

draw x_{ij} i.i.d $\sim p(x)$. ← where $p(x)$ appears

The C_n consists of the M codewords:

$C_1 = x_{11}\, x_{12} \cdots x_{1n}$

$C_2 = x_{21}\, x_{22} \cdots x_{2n}$

$\vdots$

$C_M = x_{M1}\, x_{M2} \cdots x_{Mn}$

eg. For the BSC, $p(x=0) = 0.7$, $p(x=1) = 0.3$

$n = 100$, $M = 2^{50}$

each C_i is a 100-bit string

each bit of C_i has prob 0.7 to be 0 ...

② The code C_n (i.e $C_1, C_2, \ldots, C_M$) is told to Alice & Bob

③ A message i is drawn randomly from $\{1, 2, \ldots, M\}$ by Alice

④ Alice sends C_i through $N^{\otimes n}$. (i.e $E_n(i) = C_i$.)

⑤ Bob receives output $Y^n \sim P_r(y^n | C_i) = P_r(y_1|x_{i1}) P(y_2|x_{i2}) \cdots P(y_n|x_{in})$

⑥ Bob uses the decoding map D_n:

If $\exists!\, j$ s.t. $C_j\, y^n \in A_{n,\delta}$, output j

Else "ERR".

"Joint typicality decoding"

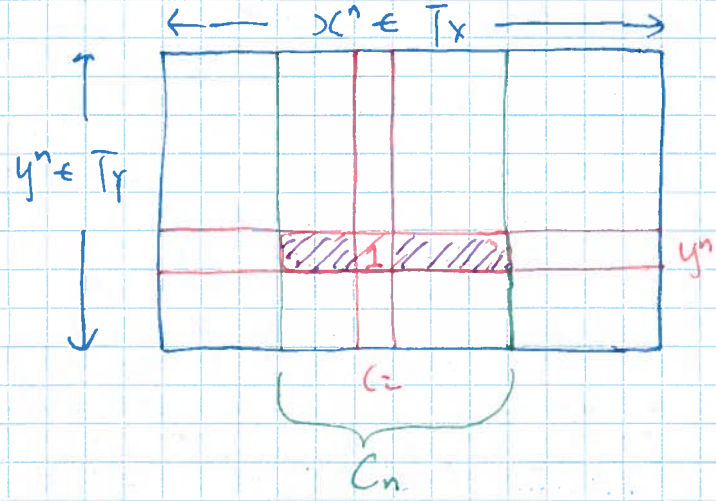
(Reminiscent of Slepian-Wolf coding, Alice has x^n, Bob has y^n, but no more communication!)

Let $T_X = T_{n,\delta}$ for $X_1 \ldots X_n$ drawn iid $\sim p(x)$

Let $T_Y = T_{n,\delta}$ for $Y_1 \ldots Y_n$ drawn iid $\sim p(y) = \sum_x p(x)\, p(y|x)$

($p(x)$: given; $p(y|x)$: from N)

Joint typicality table: (x^n, y^n)-entry $= 1$ iff $x^n y^n \in A_{n,\delta}$



Averaged over C_n & i, C_i is n iid draws $\sim p(x)$

$C_i y^n$ is n iid draws $\sim p(xy) = p(x) \cdot p(y|x)$

(a) By JAEP (1), with prob $\geq 1-\varepsilon$, $C_i y^n \in A_{n,\delta}$

(b) Assuming $C_i y^n \in A_{n,\delta}$:

Bob output $i \iff \nexists\, C_k' \neq C_i$ s.t. $C_k' y^n \in A_{n,\delta}$

(i.e. no other C_k' is in the purple area)

Note $\forall\, C_k' \neq C_i$, $C_k' \in C_n$

$\because$ C_k' indep of C_i, C_k' drawn iid $\sim p(x)$

$\because$ C_k' indep of y^n which only depends on C_i

By JAEP (3), for each $C_k' \neq C_i$

$$\Pr(C_k' y^n \in A_{n,\delta}) \leq 2^{-n(I(X;Y)-3\delta)}$$

(over C_n, $N^{\otimes n}$)

By union bound

$$\Pr\left(\exists C_k' \in S,\ C_k' \neq C_i,\ C_k' y^n \in A_{n,\delta}\right)$$

$$\leq |M| \times \max_{\substack{C_k' \in S \\ C_k' \neq C_i}} \Pr\left(C_k' y^n \in A_{n,\delta}\right)$$

$$\leq |M| \cdot 2^{-n(I(X:Y)-3\delta)}$$

Putting ⓐ & ⓑ together,

$$\mathop{\mathbb{E}}_{C_n} P_e(i) \leq \varepsilon + |M| \cdot 2^{-n(I(X:Y)-3\delta)}$$

$$\therefore \mathop{\mathbb{E}}_{C_n} \bar{P}_e(C_n) = \mathop{\mathbb{E}}_{C_n} \frac{1}{M} \sum_{i=1}^{M} P_e(i) \quad (\text{using } C_n)$$

$$= \frac{1}{M} \sum_{i=1}^{M} \mathop{\mathbb{E}}_{C_n} P_e(i)$$

$$\leq \frac{1}{M} \sum_{i=1}^{M} \left(\delta + |M| \cdot 2^{-n(I(X:Y)-3\delta)}\right)$$

$$= \varepsilon + |M| \cdot 2^{-n(I(X:Y)-3\delta)}$$

$\therefore \forall\ |M| = 2^{nR}$, $R < I(X:Y) - 3\delta - \frac{1}{n}\log\left(\frac{1}{\eta}\right)$

$\varepsilon < \eta$

We have $\mathop{\mathbb{E}}_{C_n} \bar{P}_e(C_n) \leq 2\eta$.

$\therefore \exists\ \tilde{C}_n$ s.t. $\bar{P}_e(C_n) \leq 2\eta$.

From $\tilde{C}_n$, we can get a code $\tilde{\tilde{C}}_n$ with $P_e(\tilde{\tilde{C}}_n) \leq 2 \cdot \bar{P}_e(\tilde{C}_n)$.

$\tilde{\tilde{C}}_n$ consists of $\frac{M}{2}$ codewords in $\tilde{C}_n$

with prob of error less than the median.

(ie $\tilde{\tilde{C}}_n$ = "better half" of $\tilde{C}_n$).

$\tilde{\tilde{C}}_n$ sends 1 fewer bit than $\tilde{C}_n$

Rate $\to$ by $\frac{1}{n}$, which is $I(X:Y) - 3\eta - \frac{1}{n}\log\left(\frac{1}{\eta}\right) - \frac{1}{n}$

$P_e(\tilde{\tilde{C}}_n) \leq 4\eta$.

$\therefore$ $I(X:Y)$ is an achievable rate for N.

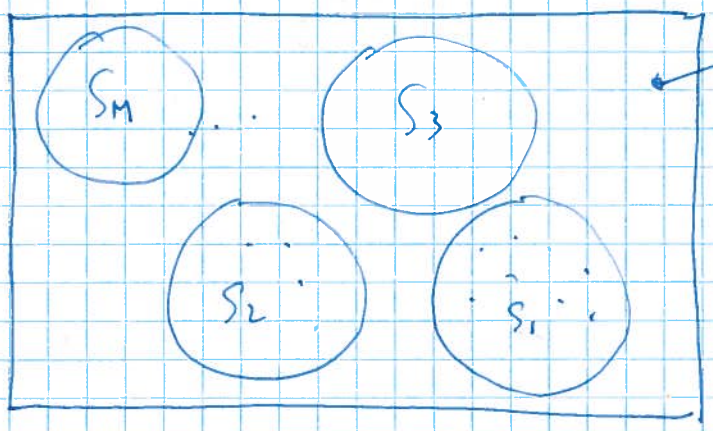
max over $p(x)$, $\max_{p(x)} I(X:Y)$ is achievable for N.

NB: Once C_n chosen $X_1 X_2 \ldots X_n$ not independent! That's why $\mathbb{E}_{C_n}$ is needed.

Techniques: Random codes, symmetry, existential proof

Expunging half of the worse codewords to boost $\bar{P}_e$ to P_e.

The geometrical picture:



where $S_i = \{y^n : (i,y^n) \in A_{n,\delta}\}$, $|S_i| \approx 2^{nH(Y|X)}$.

Intuitively, fitting $\ll 2^{nH(Y)} / 2^{nH(Y|X)} \approx 2^{nI(X:Y)}$ spheres works.

* We've bdd for a given $y^n \in S_i$, $\exists K \neq i$ s.t $y^n \in S_K$ as well.

So we've bdd the overlap between these "generalized Hamming spheres" (with prob taken into account).

The JAEP tells a lot about how these points are distributed. So we have very precise bounds.