CO781 / QIC 890:

Theory of Quantum Communication

Topic 3, part 1

Joint Typicality

Classical communication through noisy classical channel
Shannon's noisy channel coding theorem

References:

Cover & Thomas, Chapter 8

What is a noisy classical channel?

Def A classical channel N is specified by:

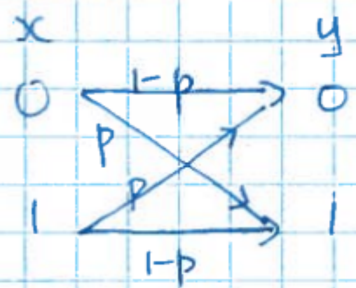
- an input alphabet Ω_X
- an output alphabet Ω_Y
- a distribution $p(y|x)$ for each $x \in X$.

eg 1 Binary symmetric channel (BSC)

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

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

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



If we choose pr(x=0) = pr(x=1) = 1/2
then we obtain the joint XY distribution:
p(00) = p(11) = (1-p)/2
p(01) = p(10) = p/2

What is a noisy classical channel?

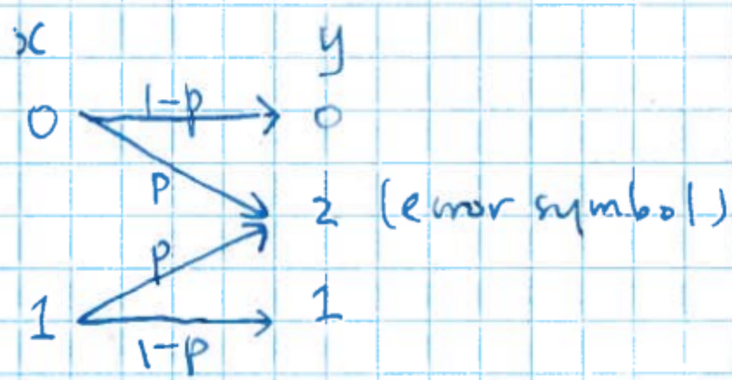
Def A classical channel N is specified by:

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

eg 2 Erasure channel

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

Input is "erased" (replaced by 2) w.p. 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 $x_1\, x_2 \ldots x_n$

the output is $y_1\, y_2 \ldots y_n$ wp $\prod_{i=1}^{n} p(y_i | x_i)$

The "use" refers to the transformation from input to output.

NB output distribution for each use is indep but NOT identical.

- Sensible channels that are out of scope:

  eg. Missing-symbol-channel (deletion errors) Opposite: insertion errors

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

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

  eg. $x_1 x_2 \ldots x_n \to x_1 x_2 \ldots x_{i-1} x_{i+1} x_i \ldots x_n$ (transposition errors)

  Symbols emerge out of order.

  eg. Burst errors

  $x_1 x_2 \ldots x_n \to x_1 x_2 \ldots x_n$

  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 $x_1, x_2 \ldots x_n$ to a code $C \subseteq \Omega_X^n$.

the subset of n-tuples Alice would ever input to the n channel uses

e.g., repetition code (n odd)

Use only 00...0 (n times) or 11....1 (n times) only to represent 1 bit.

(a) For the erasure channel, decode any bit that's not erased.
Overall prob of error p^n

(b) For the binary symmetric channel, decoding has error if more than half of the channels are erroneous, with probability

$$\sum_{k=(n+1)/2}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

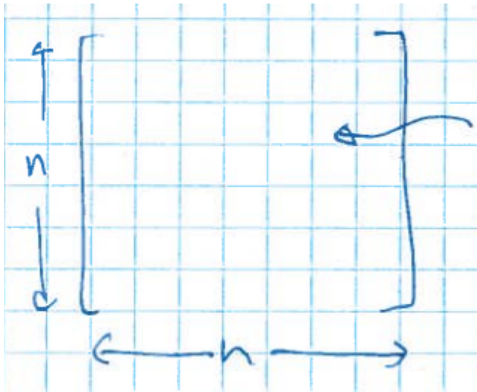
Either case, rate = 1/n.

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

the subset of n-tuples Alice would ever input to the n channel uses

e.g., Hadamard code (n multiple of 4)

Start with a Hadamard matrix of order n.



entries ± 1, mutually orthogonal rows

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

n=12

1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	1	-1	1	-1	-1	-1	1	1	1	-1	1
1	1	1	-1	1	-1	-1	-1	1	1	1	-1
1	-1	1	1	-1	1	-1	-1	-1	1	1	1
1	1	-1	1	1	-1	1	-1	-1	-1	1	1
1	1	1	-1	1	1	-1	1	-1	-1	-1	1
1	1	1	1	-1	1	1	-1	1	-1	-1	-1
1	-1	1	1	1	-1	1	1	-1	1	-1	-1
1	-1	-1	1	1	1	-1	1	1	-1	1	-1
1	-1	-1	-1	1	1	1	-1	1	1	-1	1
1	1	-1	-1	-1	1	1	1	-1	1	1	-1
1	-1	1	-1	-1	-1	1	1	1	-1	1	1

Code: the n rows of the matrix
(replace -1 by 0)

Correct decoding is at most (n/4)-1 errors with the binary sym channel.

Rate: (log n) / n.

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

the subset of n-tuples Alice would ever input to the n channel uses

e.g., binary linear code

Start with t linearly independent n-bit strings c1, c2, ..., cm.
The code consists of binary vectors v orthogonal to all ci's (mod 2).

Encode n-t bits, rate = 1-(m/n).

Error prob depends on the choice of c1, c2, ..., ct.

A real digression -- block codes are cool combinatorial objects with applications far beyond communication. e.g., cryptography.

This paper uses Reed-Solomon code to reduce the resources for testing covid by 4-6 times!

Idea: n patients, only want to run k tests for k << n.

Let 1 be attached to a sample with SARS-Cov2 DNA; very few (say, 1%)

Instead of testing individual samples, pool samples from a known subset of the patients. Result is "1" if any patient in the set is positive. (We query the "OR" function of the bits corr to the subset of patients.)

Repeat for k chosen subsets (where the ECC combinatorics come in).

Can locate all positive patients (errors) if not too many ...

Science Advances Publish Ahead of Print, published on August 21, 2020 as doi:10.1126/sciadv.abc5961

ScienceAdvances

RESEARCH ARTICLES

Cite as: N. Shental *et al.*, *Sci. Adv* 10.1126/sciadv.abc5961 (2020).

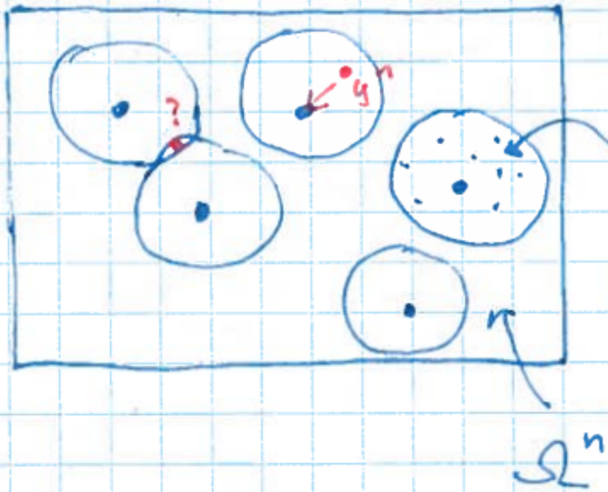
Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers

Noam Shental[1,*], Shlomia Levy[2,3]†, Vered Wuvshet[2,3]†, Shosh Skorniakov[2,3]†, Bar Shalem[4], Aner Ottolenghi[2,3], Yariv Greenshpan[2,3], Rachel Steinberg[5], Avishay Edri[2,3], Roni Gillis[6], Michal Goldhirsh[6], Khen Moscovici[6], Sinai Sachren[3], Lilach M. Friedman[2,3], Lior Nesher[5], Yonat Shemer-Avni[2,5], Angel Porgador[2,3*], Tomer Hertz[2,3,7*]

P-BEST pooling design

In our current proof-of-concept study of P-BEST, we developed a pooling scheme designed to correctly identify all positive carriers for carrier rates < 1.3%. Specifically, we pooled sets of 384 patient samples into 48 pools, each containing 48 samples. Each sample was added to six different pools. Pools were designed based on a Reed-Solomon error correcting code (*32*) which as in our previous work, proved to be robust to experimental noise, e.g., pools that fail to be amplified.

Geometric interpretation: ($\Omega = \Omega_X = \Omega_Y$ for simplicity)



$\bullet$: $x^n \in C$ codeword

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

Idea: if no more than t errors, and these spheres don't overlap, then, Bob "knows" which n-tuple (the codeword) was the input.

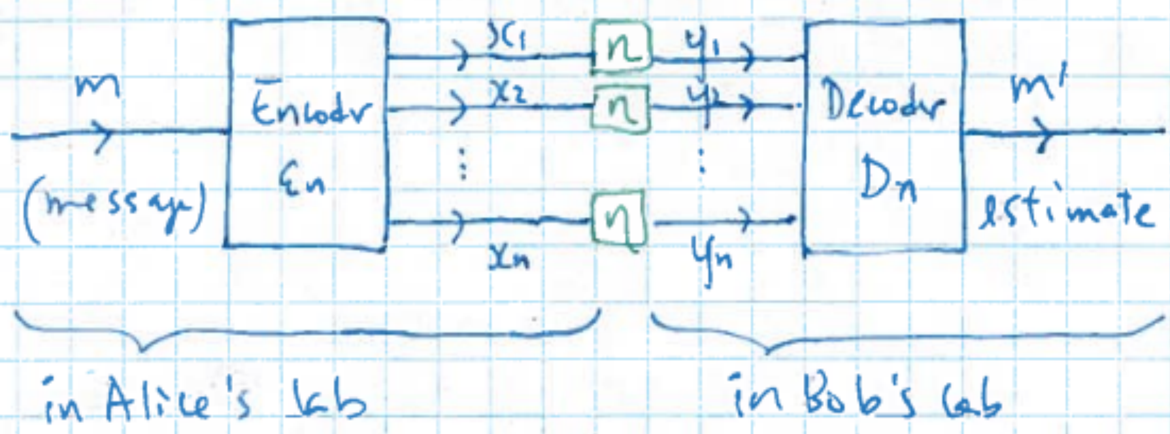
Tricky: but to suppress error, we need n to grow, this increases # errors we need to error (say, for BSC, expected # errors np).

Is it possible to have a positive rate while error vanishes with n?

NB as n grows, we want the WHOLE codeword to be output correctly, not just most of the bits.

Formal definitions before stating Shannon's noisy coding theorem:

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 $\mathcal{E}_n : \mathcal{M} \rightarrow \Omega_X^n$

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

"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 r.v with distⁿ $p(m'|m)$

$$m \longrightarrow \boxed{} \longrightarrow m'$$ is a channel

over channel noise

* Define $P_e(m) = \text{prob}(m' \neq m \mid m) = \text{prob}(D_n \circ \eta^{\otimes n} \circ \mathcal{E}_n(m) \neq m)$

$$P_e(C_n) = \max_{m \in \mathcal{M}} P_e(m)$$ 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.

Thm: Shannon's noisy coding theorem

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

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

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

NB: $C(N)$ is an asymptotic operational quantity
RHS is a "single-letter-formula"; an optimization involving one use of the channel.

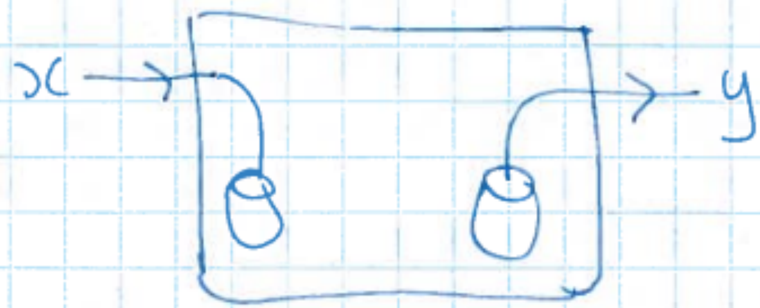
NB $p(x)$ is NOT the distribution of the input
Will see how $p(x)$ comes in.
Note the codes work in the worse case.

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

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

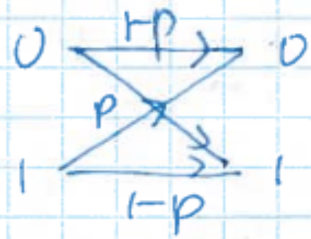
So channel discards input & draws y according to $p(y)$.

We call this channel garbage channel.



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

eg2 BSC N_p.



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

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

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

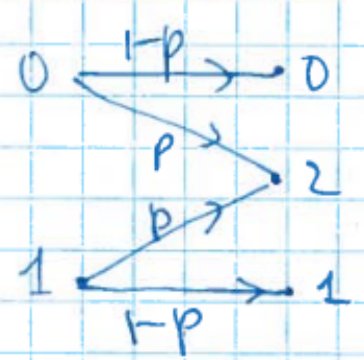
choose $p(x)$ to max $H(Y)$

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

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

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

eg3 Erasure Channel $\mathcal{E}_p$.



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

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

indep of $p(x)$

→ max this term

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

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