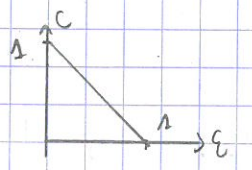
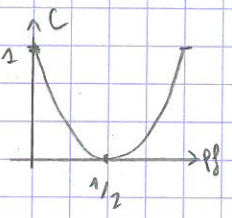
10/07/2017

" the higher, the better you can deduce input from output "

Capacity of a channel. $C = \max_{P_X} I(X;Y)$ with X input of a channel.
Y output.
P_X probability law of input.

EXERCIZES: $\begin{cases} C_{BEC} = 1-\varepsilon \\ C_{BSC} = 1-h(p_f) \end{cases}$





Encoding and decoding: message → ENCODER → encoded message → NOISY CHANNEL → corrupted encoded message → DECODER → decoded message ?

"string of bit" "longer string"

Rate of a code: $= \frac{\#\text{ bits of message}}{\#\text{ bits of encoded msg.}} < 1$ → the larger the better to reduce the cost of message transmission.

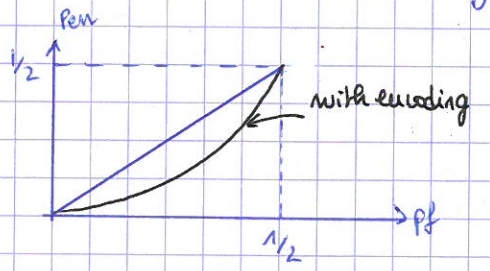
3B naive coding ≡ repetition

e.g. we repeat 3 times input. ENCODER
$0 \longrightarrow 0\,0\,0$ with the BSC $p_f < 1/2$.
$1 \longrightarrow 1\,1\,1$

reasonable DECODER "majority rule" ↳ odd number of bits
$000 \longrightarrow 0$
$111 \longrightarrow 1$
$001, 010, 100 \longrightarrow 0$
$110, 101, 011 \longrightarrow 1$

Rk: rate $= 1/3$.

How good is the naive coding ? Probability of error without encoding $= p_f$
Probability of error with 3 repetitions $= p_f^3 + 3p_f^2(1-p_f)$
↓ 3 flips ↳ $\binom{3}{2}$ flips.



→ better than no encoding but
$\begin{cases} P_{en} > 0 \text{ as soon as } p_f > 0 \\ \text{rate} = 1/3 \end{cases}$
↳ actually not that good. Can one do better ?

3C Shannon channel coding theorem

THM: There exist encoding (of growing length) with $P_{en} \xrightarrow{N\to\infty} 0$, for all rates R smaller than the capacity of the channel C ($R<C$).

So that we can now interpret the capacity as the best achievable rate with $P_{en} \to 0$.
The paradoxical statement that we could be sure of the signal despite the noise is resolved by the fact this is an asymptotic statement (thermodynamic limit).

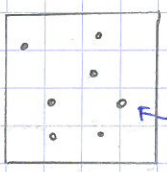
MORE FORMAL DEFINITIONS:

message	encoded message	corrupted encod. msg.	corrected encoded msg	corrected msg
$s = (s_1, \dots s_k) \in \{0,1\}^k$	$x = (x_1, \dots x_N) \in \{0,1\}^N$	$y \in X_{out}^N$ BEC $X_{out} = \{0,1,N\}$ BSC $X_{out} = \{0,1\}$	$\hat{x}(y)$	$\hat{s}(\hat{x}(y))$

↳ $N > k$

encoding: $\underline{x} = f_{\text{encoding}}(\underline{z})$

Code book $= \mathcal{C} = \{\underline{x}^{(1)}, \dots, \underline{x}^{(2^L)}\}$



exponentially many but exponentially few among available code words

channel: in $\{0,1\}^N$

intuition → put the codewords as far as possible in space
BUT! CAREFUL IN HIGH DIMENSIONS, things do not happen the same way as in 2d!) oversimplifying view.

correction: $\hat{\underline{x}}(\underline{y})$ estimation of $\underline{x}^{(\alpha)}$

decoding: $\hat{\underline{z}} = f^{-1}_{\text{encoding}}(\hat{\underline{x}}(\underline{y}))$

→ From now on we will focus on the central $\underline{x} \to \underline{y} \to \hat{\underline{x}}(\underline{y})$. If the 2^L $\underline{z}$ are uniform at random, the $\underline{x}$ are also u.a.r.

DECODING AS AN INFERENCE PROBLEM:

$\underline{X}$ random codeword, $P_{\underline{X}}(\underline{x}) = 1/2^L \; \mathbb{1}(\underline{x} \in \mathcal{C})$

$\underline{Y}$ random output $P_{\underline{Y}|\underline{X}}(\,|\,) = \prod_{i=1}^{N} Q(y_i|x_i)$ ← over the bits indep.

BEC: $\begin{cases} Q(0|0) = 1-\varepsilon \\ Q(*|0) = \varepsilon \\ Q(1|0) = 0 \end{cases}$

using Bayes theorem: $P_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = P_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) \dfrac{P_{\underline{X}}(\underline{x})}{P_{\underline{Y}}(\underline{y})}$ ← prior proba on the signal

$$= \prod_i Q(y_i|x_i) \frac{1}{2^L} \mathbb{1}(\underline{x} \in \mathcal{C}) \frac{1}{P_{\underline{Y}}(\underline{y})}$$

One keeps only the factors function of $\underline{x}$ and lets the rest into the partition $Z(\underline{y})$.

$$\equiv \frac{1}{Z(\underline{y})} \mathbb{1}(\underline{x} \in \mathcal{C}) \prod_{i=1}^{N} Q(y_i|x_i)$$ POSTERIOR PROBABILITY

Rk: Having all the distribution of the decoding given the observation implies that we have 2 ways of decoding:

Maximal a posteriori.

* $\hat{\underline{x}}(\underline{y}) = \underset{\underline{x}}{\operatorname{argmax}} \; P_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y})$ block MAP → min $P(\hat{\underline{X}} \neq \underline{X})$

* $\hat{x}_i(\underline{y}) = \underset{x_i}{\operatorname{argmax}} \; P_{X_i|\underline{Y}}(x_i|\underline{y})$ symbol MAP decoding → min $\mathbb{E}(d(\hat{\underline{X}}, \underline{X}))$

EXAMPLE WITH THE BEC: $\underline{y} = (0, 1, 0, 0, *, *, 0 \dots *, 0)$ (the $*$ are the erasures)

$$P_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{1}{|\mathcal{C} \cap B(\underline{y})|} \mathbb{1}(\mathcal{C} \cap B(\underline{y}))$$

$B(\underline{y}) = \{\underline{x} \in \{0,1\}^N \text{ s.t. } y_i \in \{0,1\} \Rightarrow x_i = y_i\}$

intersection codebook and ball

uniform probability over the messages matching the revealed bits in received message.

How should we construct code books? Let's start with a simple construction proposed by Shannon.

SHANNON RANDOM CODE ENSEMBLE

→ connection to the REM!

$\mathcal{C} = \{\underline{x}^{(1)}, \dots \underline{x}^{(2^L)}\}$ with $\underline{x}^{(\alpha)} = \{x_1^{(\alpha)}, \dots x_N^{(\alpha)}\}$ → choose the $N \times 2^L$ $x_i^{(\alpha)}$ iid 0,1 with probability $1/2$.

Rk: proba that $\underline{x}^{(\alpha)} = \underline{x}^{(\beta)}$ for $\alpha \neq \beta$ (non-injective) $\xrightarrow[\text{L} \to +\infty]{N \to +\infty} 0$ $R = \frac{L}{N}$ fixed.

Bin$(N, 1-\varepsilon)$ fluctuations

ANALYSIS ON THE BEC:

Assume w.l.o.g that $\underline{x}^{(1)}$ has been sent: $\underline{y} = \{\underline{x}^{(1)}$ on $\sim N(1-\varepsilon) \pm O(\sqrt{N})$ correctly transmitted
$*$ on other bits $\sim N\varepsilon$

$W \overset{d}{=} \text{Bin}\left(2^L - 1, \left(\frac{1}{2}\right)^{N(1-\varepsilon)}\right)$ "number of confusing code word".

(in distribution; $2^L - 1$: total # of available code words; $\left(\frac{1}{2}\right)^{N(1-\varepsilon)}$: proba 2 ≠ codewords agree on a bit)

$\Rightarrow \mathbb{E}(W) = 2^{N(R-1+\varepsilon)} \to 0$ if $\varepsilon < 1 - R$ $\Rightarrow$ $P_r[W \neq 0] \to 0$ - (↳ integer!)

First moment / Markov inequality for strictly positive codeword -

So that no confusion if $R < 1 - \varepsilon = C_{BEC}$ → Shannon theorem! ∎

Even the random codes achieve capacity - (Here we considered averages for codewords etc...)

HOMEWORK: proof for the BSC (text on website) -

Rk we proove that there exist such codes, but we did not show that we could not do better than this bound - if $R > C \Rightarrow P_{err} > 0$ coming from $H(\underline{X}|\underline{Y}) > 0$, Fano inequality -

⚠ In practice, encoding and decoding the Random Code ensemble takes exponential time and memory - $\mathcal{C}$ has to be described by $N 2^L$ bits. The encoding is easy by looking up the table - But the decoding is NP hard as one need to look for the corresponding match in the table - The randomness hinders any compression, we can only do exhaustive (es) search in the table -

4 LOW DENSITY PARITY CHECK CODES (LDPC). → putting some structure ;)

4A Linear codes

- $\{0,1\}^N$ is a linear space over $\mathbb{Z}_2 = \{0,1\}$ "scalars", addition mod 2, multiplication
- The codebook $\mathcal{C} \subset \{0,1\}^N$ is said to be a linear code if $\mathcal{C}$ is a linear subspace of $\{0,1\}^N$.
  - $\underline{x} + \underline{y} = (x_1 + y_1 \text{ mod } 2, x_2 + y_2, \dots x_N + y_N) \in \mathcal{C}$
  - $\underline{x} + \underline{x} = \underline{0} \in \mathcal{C}$ (the origin belongs to the linear subspace).
  - ↳ $\underline{0}$ is always a codeword of a linear code
  - All codewords are equivalent (can always make gauge transformation to change position of origin).

12/07/2017 RECALL THE IDEA:

$\#\mathcal{C} \ll \#\{0,1\}^N$: if the codewords are sufficiently far from one another, we should be able to identify them despite of the noise corruption by the channel -
$2^L \qquad 2^N$

$R = \frac{L}{N} < 1$

↳ the codebook can be described by a set of M linear equation, or equivalently as the Kernel of an operator: $\mathcal{C} = \ker H = \{x \in \{0,1\}^N : H\underline{x} = \underline{0}\}$. $H = \begin{bmatrix} 0\,0\,1\,0\,\dots \\ 0\,0\,0\,1\,\dots \\ \dots \end{bmatrix}$ (M rows, N columns)

↳ one line of $H\underline{x} = 0 \Rightarrow x_3 + x_{14} + x_{25} = 0 =$ nb of 1's among $\{x_3, x_{14}, x_{25}\}$ must be even → hence the name PARITY CHECKS

prop: $\dim \mathcal{C} =$ nb of linearly independent codewords

$|\mathcal{C}| = 2^{\dim \mathcal{C}}$

number of codewords = number of solutions

↳ suppose M equations in $H\underline{x} = \underline{0}$ are independent, then $|\mathcal{C}| = 2^{N-M}$ -

general result of linear algebra: $N = \underbrace{\dim \ker H}_{=L} + \underbrace{\text{rk}(H)}_{=M \text{ if } H \text{ is full-rank}} \to R = \frac{N-M}{N} = 1 - \frac{M}{N}$ $M < N$.