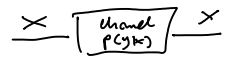
3 - Channel capacity

7 października 2010
16:06

3.1 Mutual information

X, Y two random variables



how much do we learn about X once we know Y
intuitively this is how much information channel has transmitted

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

$I(X;Y) \geq 0$ ($=0$ for independent variables)

Def: conditional mutual information

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

Fact
(Chain rule for I:)

$$I(X_1, X_2 : Y) = I(X_1; Y) + I(X_2 : Y | X_1)$$

Proof:

$L = H(X_1X_2) + H(Y) - H(X_1, X_2, Y)$

$R = H(X_1) + H(Y) - H(X_1, Y) + H(X_2|X_1) + H(Y|X_1) - H(X_2, Y|X_1)$

$H(X_2, Y|X_1) = H(X_2|YX_1) + H(Y|X_1)$

$-H(X_1, X_2, Y) = -H(X_1, Y) - H(X_2|YX_1)$

$0 = 0$ ∎

3.2 Data processing inequality

$X \xrightarrow{p(y|x)} Y \xrightarrow{p(z|y)} Z$ } Markov chain

$p(x, y, z) = p(z|y)\, p(y|x)\, p(x)$

a) $I(X:Y) \geq I(X:Z)$ } alternatively $H(X|Y) \leq H(X|Z)$

b) (i) $I(Y:Z) \geq I(X:Z)$ } $H(Z|Y) \leq H(Z|X)$

Proof:

c) $I(X:YZ) = I(Y:X) + I(Z:X|Y)$

$= I(Z:X) + I(Y:X|Z)$

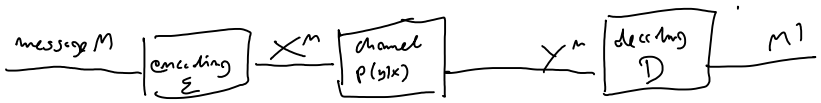
$I(Z:X|Y) = H(Z|Y) - H(Z|XY) = 0$

↑ not relevant

$I(Y:X) = I(Z:X) + I(Y:X|Z) \Rightarrow I(Y:X) \geq I(Z:X)$ ∎

(ii) analogously

3.3 Shannon channel theorem (the greatest achievement of information theory)



code: $(|M|, n)$

Max error probability $\lambda_n = \max_{m \in M} \Pr\left[D(Y^n) \neq m \mid X^n = E(m) \right] =$

$= \sum_{y^n} p(y^n \mid x^n(m)) \cdot \mathrm{If}\left[D(y^n) \neq m, 1, 0 \right]$

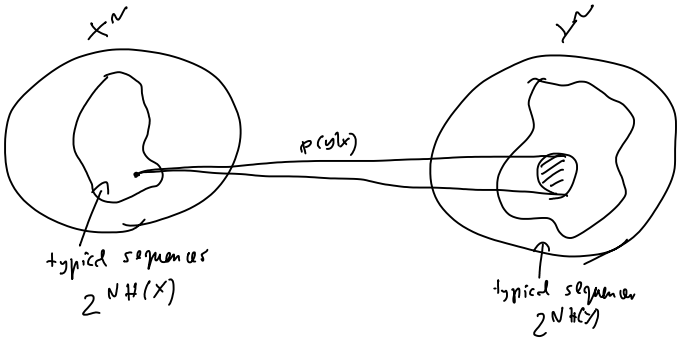
Definition Rate of the code $R = \frac{\log |M|}{n}$

Definition A rate R is achievable if there exist a sequence of $(2^{nR}, n)$ codes where maximal probability

of error $\lambda_n \xrightarrow{n\to\infty} 0$.

Theorem: R is achievable iff $R \le C = \max_{p_X} I(X:Y)$

Intuition:



one typsequence in X^N will be transformed to

$2^{NH(Y|X)}$ =

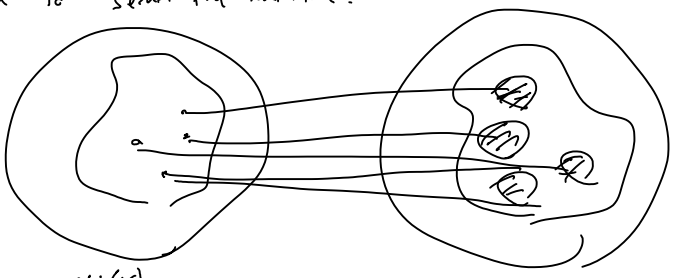
$Np(x)$ very typ. x contains probability mass

$2^{Np(x)H(Y|x)}$ ungiven adding with $Np(x)$

to some other knowledge x. Many vice:

$2^{\sum_x Np(x)H(Y|x)} = 2^{NH(Y|X)}$ typ sequences

how many different input sequences can be used to send information:



$$\frac{2^{NH(Y)}}{2^{NH(Y|X)}} = 2^{NI(X:Y)}$$

Mutual information tells us about channel capacity

Strictly:

. For every rate $R < C = \max_{p_X} I(X:Y)$ there exists a

(1) coding $(2^{nR}, n)$ with probability of error λ arbitrarily small for $n \to \infty$.
number of messages / └ number of symbols transmitted

(2) Conversely if $(2^{nR}, n)$ exists $\Rightarrow$ $R \le C$

Definition: Jointly typical sequences T_ε^n

(x^n, y^n) are jointly ε-typical with respect to $p(x,y)$ iff

- $\left| -\frac{1}{n} \log p(x^n) - H(X) \right| \le \varepsilon$
- $\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| \le \varepsilon$
- $\left| -\frac{1}{n} \log p(x^n, y^n) - H(X,Y) \right| \le \varepsilon$ $\quad p(x^n, y^n) = \prod_i p(x_i, y_i)$

Properties also asymptotically for large n

$P[(x^n, y^n) \in T^n] \ge 1 - \varepsilon$

(i) $\Pr[(x^n,y^n)\in T_\varepsilon^n] \geq 1-\delta$

(ii) $(1-\delta)\,2^{n(H(X,Y)-\varepsilon)} \leq |T_\varepsilon^n| \leq 2^{n(H(X,Y)+\varepsilon)}$

(iii) If $\tilde{X}^n, \tilde{Y}^n$ are independent variables with the same marginals as X^n Y^n: $p(\tilde{x}^n,\tilde{y}^n)=p(x^n)p(y^n)$ then

$$(1-\delta)\,2^{-n(I(X:Y)+3\varepsilon)} \leq \Pr\left((\tilde{X}^n,\tilde{Y}^n)\in T_\varepsilon^n\right) \leq 2^{-n(I(X:Y)-3\varepsilon)}$$

Proofs

(i) By law of large numbers for sufficiently large n we have

$\Pr\left[\left|-\frac{1}{n}\log p(x^n) - H(X)\right| > \varepsilon\right] \leq \frac{\delta}{3}$

$\Pr\left[\left|-\frac{1}{n}\log p(y^n) - H(Y)\right| > \varepsilon\right] \leq \frac{\delta}{3}$ $\Rightarrow \Pr\left[(x^n,y^n)\notin T_\varepsilon^n\right] \leq \delta$

$\Pr\left[\left|-\frac{1}{n}\log p(x^n,y^n) - H(X,Y)\right| > \varepsilon\right] \leq \frac{\delta}{3}$

(ii) $p(x^n,y^n) \geq 2^{-n(H(X,Y)+\varepsilon)}$ $\quad \sum_{x^n,y^n\in T_\varepsilon^n} p(x^n,y^n) \leq 1$

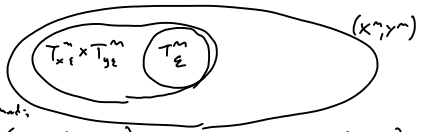
$|T_\varepsilon^n| \cdot 2^{-n(H(X,Y)+\varepsilon)} \leq 1$ $\quad |T_\varepsilon^n| \leq 2^{n(H(X,Y)+\varepsilon)}$

$p(x^n,y^n) \leq 2^{-n(H(X,Y)-\varepsilon)}$

$$1 = \sum_{(x^n,y^n)} p(x^n,y^n) = \sum_{(x^n,y^n)\in T_\varepsilon^n} p(x^n,y^n) + \sum_{(x^n,y^n)\notin T_\varepsilon^n} \leq |T_\varepsilon^n|\,2^{-n(H(X,Y)-\varepsilon)} + \delta$$

$$|T_\varepsilon^n| \geq (1-\delta)\,2^{n(H(X,Y)-\varepsilon)}$$

(iii)



upper bound:

$$\Pr\left((\tilde{x}^n,\tilde{y}^n)\in T_\varepsilon^n\right) = \sum_{(x^n,y^n)\in T_\varepsilon^n} p(x^n)p(y^n)$$

$$\leq \underbrace{2^{-n(H(X)-\varepsilon)-n(H(Y)-\varepsilon)}}_{\text{upper bound on probability of } p(x^n),p(y^n)} \cdot \underbrace{2^{n(H(X,Y)+\varepsilon)}}_{\text{upper bound on } |T_\varepsilon^n|} = 2^{-n(I(X:Y)-3\varepsilon)}$$

lower bound:

$$\Pr\left((\tilde{x}^n,\tilde{y}^n)\in T_\varepsilon^n\right) = \sum_{(x^n,y^n)\in T_\varepsilon^n} p(x^n)p(y^n) \geq (1-\delta)\,2^{n(H(X,Y)-\varepsilon)} \cdot 2^{-n(H(X)+\varepsilon)}\,2^{-n(H(Y)+\varepsilon)}$$

$$= (1-\delta)\,2^{-n(I(X:Y)+3\varepsilon)}$$

Proof of channel coding theorem

(1) Let us fix $p(x)$. We generate 2^{nR} random sequences x^n $\rightarrow$ we obtain a certain $\{2^{nR}, n\}$ code $\mathcal{C}$. Probability to generate a particular code

$$p(\mathcal{C}) = \prod_{m=1}^{2^{nR}} \prod_{j=1}^{n} p(x_j(m))$$

We choose a message m $\longrightarrow$ $x^n(m) = (x_1(m),\dots,x_n(m))$ (encoding)

receivers receives y^n with probability $p(y^n|x^n(m)) = \prod_{i=1}^{n} p(y_i|x_i(m))$

He decodes a message m' provided such that $(x^n(m'), y^n)$ is a jointly typical sequence and there is no other w for which $(x^n(w), y^n)$ is typical, otherwise an error is declared.

What is the probability of error?

probability of error averaged over all choice of codes for a single message m: (by symmetry nothing depends on m)

message m: (by symmetry nothing depends on m)

$$P_e(m) = pr\left[(y^n, x^n(m)) \text{ is not jointly typical (a)} \right] + pr\left[\exists_{m' \neq m} \left((y^n, x^n(m')) \text{ is jointly typical (b)} \wedge (y^n, x^n(m)) \text{ is jointly typical} \right) \right]$$

(a) $pr \leq \varepsilon$

(b) y^n : $x^n(m')$ by construction of codes are independent variable

$pr((y^n x^n(m')) \text{ is jointly typical}) \leq 2^{-n(I(X:Y)-3\varepsilon)}$

We have 2^{nR} messages so:

$$pr\left[\exists_{m' \neq m} (y^n, x^n(m')) \text{ is jointly typical} \right] \leq 2^{nR} \cdot 2^{-n(I(X:Y)-3\varepsilon)}$$

Finally:

$$P_e \leq \varepsilon + 2^{n(R - I(X:Y) + 3\varepsilon)}$$

so if we choose $R < I(X:Y) + 3\varepsilon$ we get expanential decrease of P_e (1)

(2) Assume code $(2^{nR}, n)$ is error free:

for any message $\sum_{y^n} p(y^n | x^n(m)) \cdot \mathbb{1}[D(y^n) \neq m, 1, 0] = 0$

This implies: $H(M | Y^n) = 0$ y^n determines m uniquely.

Assume M is uniformly distributed:

$$nR = H(M) = I(M : Y^n) + \underbrace{H(M | Y^n)}_{0} \leq$$

data processing ineq $\leq I(X^n : Y^n) \leq$ see below $\sum_i I(X_i : Y_i) \leq n \cdot C$

$I(X^n : Y^n)$ $p(y^n, x^n) = p(y^n | x^n) \cdot p(x^n) = \prod_i p(y_i | x_i) \cdot p(x^n)$

$$H(X^n, Y^n) = -\sum_{x^n y^n} p(y^n, x^n) \log p(y^n, x^n) = -\sum_{x^n y^n} \prod_i p(y_i|x_i)\, p(x^n) \left(\log p(x^n) + \sum_i \log p(y_i|x_i) \right) =$$

$$= H(X^n) + \sum_{x^n} \sum_i H(Y_i | X = x_i)\, p(x^n) = H(X^n) + \sum_i H(Y_i | X_i)$$

$$I(X^n : Y^n) = H(X^n) + H(Y^n) - H(X^n, Y^n) = H(Y^n) - \sum_i H(Y_i | X_i) \leq$$

$$\leq \sum_i H(Y_i) - \sum_i H(Y_i | X_i) = \sum_i I(Y_i | X_i)$$

So we have $R \leq C$

- Now allowing for a small probability of error: ε

$$M \to X^n \to Y^n$$

Lemma Fano inequality

X, Y - two random variables.

What is the probability of guessing X correctly, knowing Y

$p(x)$, $p(y|x)$, obtaining y we calculate $g(y) = \tilde{x}$

We can think of it as a Markov chain $X \to Y \to \tilde{X}$

Probability of error $P_e = P(X \neq \tilde{X})$

Th: $H(P_e) + P_e \log(|\mathcal{X}|-1) \geq H(X|Y)$

Proof

Let E be random variable $= \begin{cases} 0 & X = \tilde{X} \\ 1 & X \neq \tilde{X} \end{cases}$

by chain rule:

$$H(EX|Y) = H(E|Y) + H(X|EY)$$
$$= H(X|Y) + \underbrace{H(E|XY)}_{0}$$ since $g(y)$ is deterministic

$$H(E|Y) \leq H(E) = H(P_e)$$
$$H(X|E,Y) = \sum_{e=0,y} p(0,y) \underbrace{H(X|e=0,y)}_{0} + \sum_{e=1,y} p(1,y) H(X|e=1,y) \leq P_e \log(|\mathcal{X}|-1)$$
$$H(P_e) + P_e \log(|\mathcal{X}|-1) \geq H(E|Y) + H(X|EY) = H(X|Y)$$

$\Downarrow$

$$1 + P_e \log|\mathcal{X}| \geq H(X|Y)$$
$$P_e \geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|}$$

Using Fano inequality:

$$\varepsilon \geq \frac{H(M|Y^n) - 1}{nR}$$
$$\varepsilon nR + 1 \geq H(M|Y^n)$$

By data processing inequality we have

$$\varepsilon nR + 1 \geq H(X^n|Y^n)$$

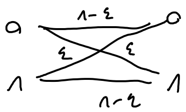
– M be a uniform distribution

$$nR = H(M) = I(M:Y^n) + H(M|Y^n) \leq I(X^n(M), Y^n) + \varepsilon nR + 1$$
$$< nC + \varepsilon nR + 1$$
$$\leq C + \varepsilon R + \frac{1}{n} \quad \xrightarrow{n \to \infty} \quad R \leq C$$

shannon coding not really practical

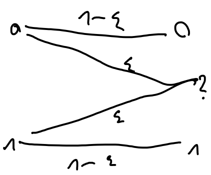
3.4 Examples

3.4.1 Binary symmetric channel ($\varepsilon = 10\%$)



$C = 1 - H(\varepsilon)$

3.4.2 Binary erasure channel



$C = 1 - \varepsilon$

3.5 Hamming codes (practical but suboptimal code)

- the simplest codes: repetition codes + majority voting (not very useful)
- better, use parity checks
  - $n-1$ bits of information, add 1 bit such that parity is zero
  - allows to detect odd number of errors, but cannot correct
- Hamming code $[7, 4]$, 7 bit code with 4 bit message correcting one error

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

all non-zero 3 bit columns

codespace $\mathcal{C} = \ker H$, $\dim \mathcal{C} = 4$

$\mathcal{C} =$

Table 8.1. The Hamming G,4) Code

0000000	0001111	0010110	0011001
0100101	0101010	0110011	0111100
1000011	1001100	1010101	1011010
1100110	1101001	1110000	1111111

← all codewords except zero one have at least 3 1s (because columns of H are independent)

All codewords differ in at least 3 places – minimum weight.

$Hc = 0$ if we have an error $r = c + e_i$ ← error in i-th place

$Hr = Hc + He_i =$ i-th column of H (we detect error and know where it happened)

$[7, 4, 3]$ – Hamming code

block length ↑, information bits ↑, ↳ minimum weight

- General Hamming code correcting 1 error

$$H = L\Big[\quad\Big\}\ 2^L - 1 \qquad [2^L - 1,\ 2^L - 1 - L,\ 3]$$

- For correcting more errors ex Reed-Solomon codes