8 - Shannon

18 listopada 2014
11:47

Shannon Entropy & Mutual information

- Shannon entropy: $H(X) = -\sum_{x} p(x) \log_2 p(x)$
  - miara przypadkowości zmiennej losowej X
  - $X = 0,1$ $p(0) = \frac{1}{2}$ $p(1) = \frac{1}{2}$ $H(X) = 1$
  - $p(0) = 1$ $p(1) = 0$ $H(X) = 0$

  Jak dobrze można skompresować dane.

Example:

a, b, c, d $p(a) = \frac{1}{2}$ $p(b) = \frac{1}{4}$ $p(c) = p(d) = \frac{1}{8}$

How to encode to use on average the smallest number of bits

$a = 0$
$b = 10$ on average 1,75 bits
$c = 110$
$d = 111$

$H(X) = 1,75$.

If we chose a different encoding e.g. $a = 00, b = 01, c = 10, d = 11$ we would see that we would use 2 thus 1 - not optimal

Intuition:

If we have random variable X recited N times then for large N we will always have sequences in which symbol x appears $\simeq N p(x)$ times - typical sequence.

- Law of large numbers.

probability of a given typical sequence

$$p_{typical}^{N} = \prod_{x} p(x)^{N p(x)} = 2^{N \sum_x p(x) \log_2 p(x)}$$

$$N_{\substack{typical \\ sequences}} = \frac{1}{p_{typical}^{N}} = 2^{-N \sum_x p(x) \log_2 p(x)} = 2^{N H(X)}$$

So we need $N H(X)$ bits to label all sequences

In this sense $H(x)$ is the compression rate.

Shannon source coding theorem.

- Shannon mutual information

X ——— channel ——— Y

$p(y|x)$

$p(x)$
└ probability distribution for input symbols

- conditional entropy:

$$H(Y|x) = -\sum_y p(y|x) \log p(y|x)$$

on average:

$$H(Y|X) = \sum_x p(x) H(Y|x) = -\sum_{xy} p(x) p(y|x) \log p(y|x)$$

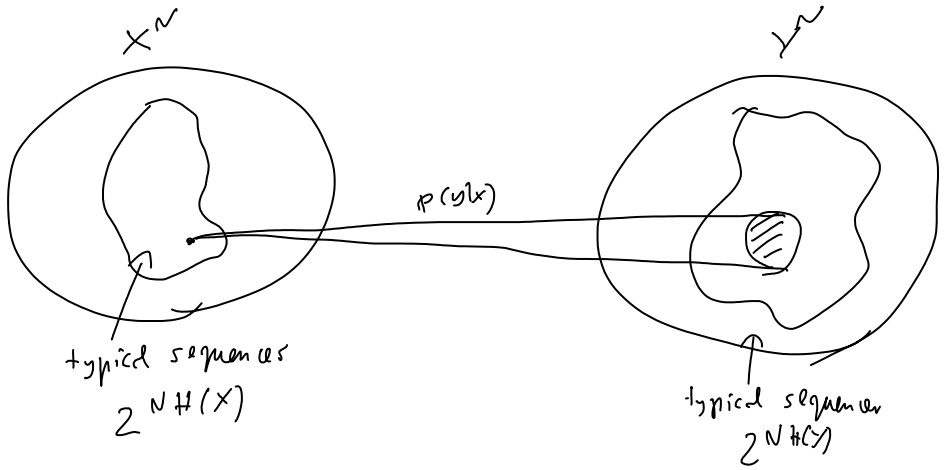
how random is Y if we know X

- mutual information
(how much do we learn about Y once we learn X)

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

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)$$

Intuition



one typ.seq. in X^N will be transformed to

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

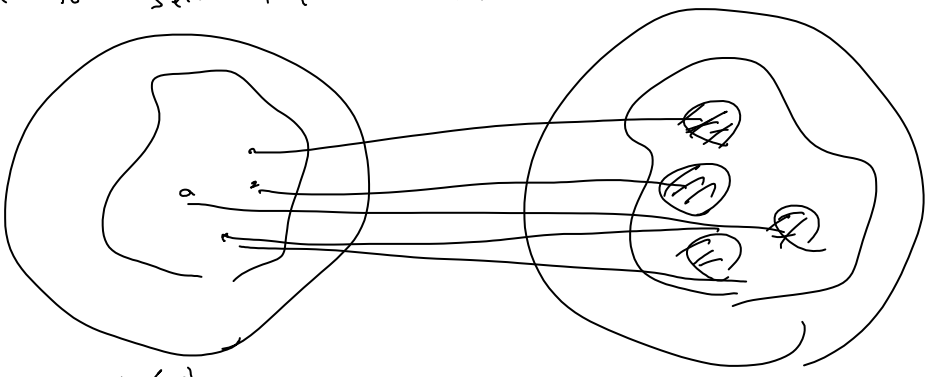
$Np(x)$ wzy wystąpi x zostanie pomnożony na

$2^{Np(x) H(Y|x)}$ ciągów długości $Np(x)$

to samo dla każdego x. Mamy więc:

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

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



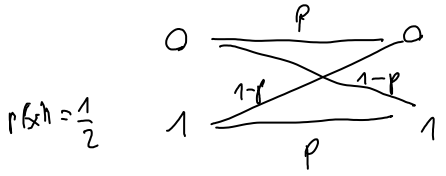
$$\frac{2^{NH(Y)}}{2^{NH(Y|X)}} = 2^{NI(X:Y)}$$

Mutual information tells us about channel capacity

- Shanno channel theorem:

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

Example:

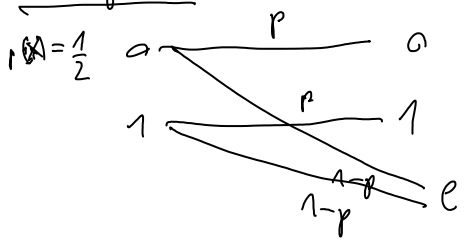


$$h(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

$$I(X:Y) = 1 - h(p)$$

- $p = 100\%$ $\quad I(X:Y) = 1$
- $p = 50\%$ $\quad I(X:Y) = 0$
- for $p = 90\%$ $\quad I(X:Y) = 0{,}531$

Example:



$$H(Y) = -\frac{1}{2}p\log_2\frac{p}{2} - \frac{1}{2}p\log_2\frac{p}{2} - (1-p)\log(1-p)$$

$$H(Y|X) = 2\cdot\frac{1}{2}\left(-p\log p - (1-p)\log(1-p)\right)$$

$$I(X:Y) = -p\log\frac{p}{2} + p\log p = p$$

Application of Shannon theorem to QKD

After the sifting stage we can think that three parties have correlated strings of N bit

A 1101c1c1c10

B 11[1]1 a1a 1 a [a] a error

E 1[a]a1 [1]1a 1 1[1] a [a] same inf on the key

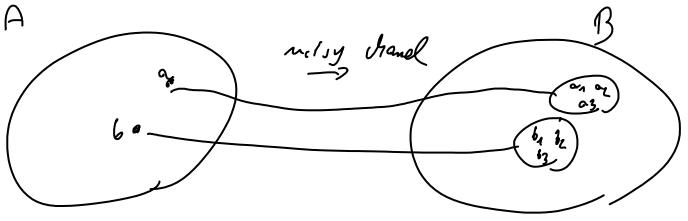
A and B want to correct errors and make sure E know nothing

Assumption: we can treat each realization as independent: we have N independent realization of three random variables $\{A,B,E\}^N$.

Error - correction

Intuition based on Shannon channel theorem

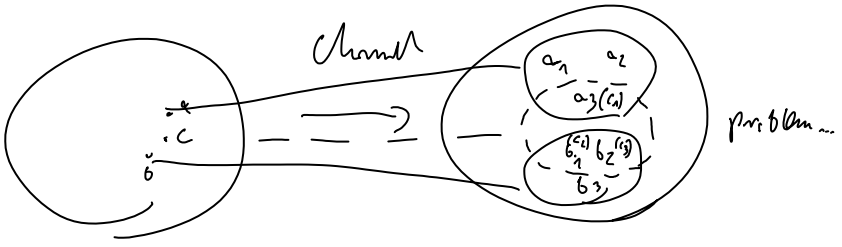
Assumption, errors are independent $\Rightarrow$ A and B



In Shana we used lower codewords to avoid ambiguity - „Error correction before".

In QKD we have to do „Error Correction after.

B can distinguish points which are more distant then a „sphA"



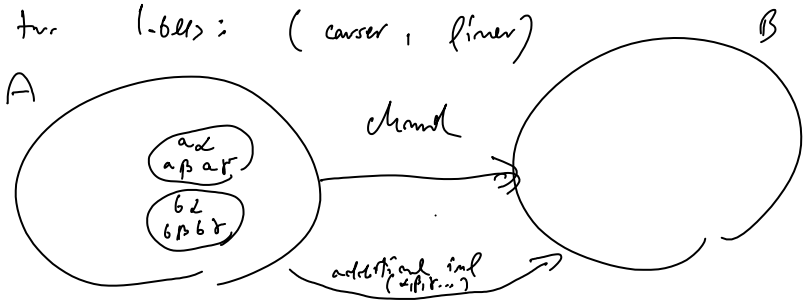
B can distinguish between codewords which are sufficiently separated.

There are $2^{NH(Y)}$ typ seq at B

from which he could distinguish $2^{NH(Y|X)}$

We could label A sequences with

We could label A sequences with two labels: (coarser, finer)



If A provides B with the finer label he can decode himself the coarser label

So A needs to additionally send:

$$N(H(A) - I(A:B)) = N(1 - I(A:B))$$

to give B a chance to decode without errors.

We have to remember that this amount of information is also available to E.

In practice

- Interactive error-correction protocol (1992, Bennett et al)

Iteration:

- A and B apply random permutation
- A and B divide their N bits in blocks of length n

$$N = k \cdot n$$

(the length of the block should be such that it is not very probable that there are more than 1 error)

- They check parities of bits in each subblock if it does not agree → bisection
- If all parity errors were corrected they repeat the iteration with larger block size ...

- Maximal sensible size of the block is $k = \frac{N}{2}$
If a number of subsequent checks (e.g. 20) yield no errors it is almost certain $(1 - 2^{-20})$ that no errors are left.

Privacy amplification

Assume A & B have corrected all the errors. If the revealed info allowed E to correct her error we can do nothing.

So we need the condition:

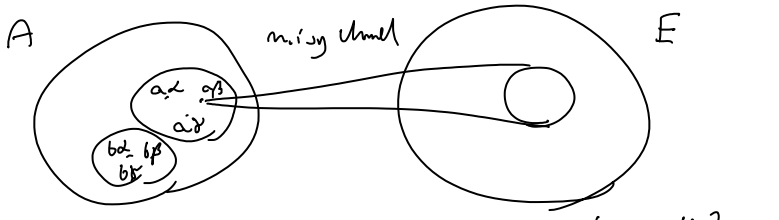
$$I(A:E) < I(A:B)$$

Assuming this is the case what can we do?

We can shrink the sequence so that E errors plague the first sequence in a way that $I(A':E') \approx 0$.

We denote A', B', E' a situation after E-C.

(we can look at it as "anti error correction")



we know that E is able to distinguish the latin label but not the greek

So A and B take only "greek" labels as they secret key on which E has no knowledge

How long is the string:

$$N\left(H(A') - \overset{H(A)}{I(A':E')}\right)$$

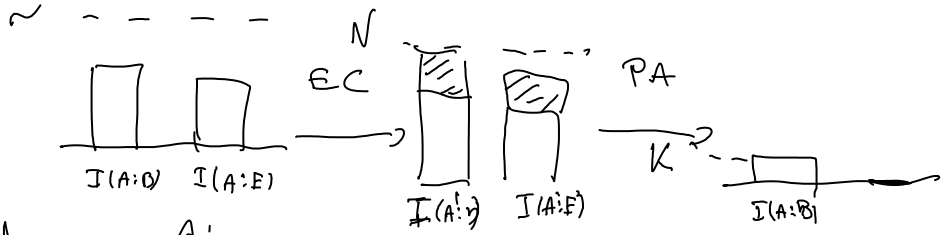
but $I(A':E') \approx I(A:E) + \left(H(A) - I(A:B)\right)$

So [illegible]

so finally: the length of the secure key could in theory be:

$$K = N\left(I(A;B) - I(A:E) \right)$$

(Csiszar-Körner theorem)



In practice

- a simple example that does the job

Choose a family of $K \times N$ random 0-1 matrices, and apply one of them:

$$K \left\{ \begin{bmatrix} \\ \end{bmatrix} = \begin{bmatrix} M_{K' \times N} \end{bmatrix} \begin{bmatrix} \\ \\ \end{bmatrix} \right\} N$$

(example of a hashing function)

Looking for the optimal attack in QKD

In order to implement QKD we need to know optimal attack for a given QBER so we know how much do shrink new key:

$$K = N\left(I(A:B) - I^{\text{optimal}}(A:E) \right)$$

$I(A:B) = 1 - h(QBER)$, $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$

np. w BB84

$|\psi\rangle_A = |0\rangle, |1\rangle, |+\rangle, |-\rangle$

max $I(A:E)$ with constraints on QBER = q.
$\{U, M\}$