Information Theory

- An information source generates messages from a message set
- Message set can be
  - descrete
  - continuous
- A sequence of messages at the output of the source forms a discrete / continuous random process.
- If messages are statisticaly independent then we say that the source is without memory. Otherwise, the source has memory.

- Definitions of information:

  How much information does a message carries?

  1) – If there is only one message, we don't have to send it, the recipient knows what to expect.

  2) – The recipient received a message but he/she did not understand it. (e.g. the message is in a different language)

  3) – The recipient understands the message but this understanding is not of any use.

- Information content can be considered at different levels

- sintax : - there is some ambiguity, uncertainty before the message was received.
  - the meaning is irrelevant.
  - the problem is to deliver a message correctly.

- semantic : - the recepiant must understand the message
  e.g. automatic translation
  -

- pragmatic : - value of information to the user

Meassures of information

- Consider a discrete memoryless source
- Let the set of messages is $S = \{S_1, S_2, \dots, S_m\}$ (or symbols)
- and their probabilities are $P_1, P_2, \dots P_m$

$$Q(S_i) = \log \frac{1}{P_i}$$ - amount of information (self information)

- Let consider two independent symbols S_i, S_j

$$Q(S_i, S_j) = \log \frac{1}{P(S_i, S_j)} = \log \frac{1}{P(S_i)P(S_j)} = \log \frac{1}{P(S_i)} + \log \frac{1}{P(S_j)}$$

The base of the algorithm is arbitrary

eg.	base	unit
	2	Shannon
	10	Hartley
	e	nat

Some other deffinitions:

Viewer : - surprise index $SI(S_i) = \frac{\sum_{i=1}^{m} p^2(S_i)}{p(S_i)}$ - example chess board

Fisher :

$$I = \int_{-\infty}^{+\infty} \frac{\partial \ln(p(x))}{\partial m} p(x) dx$$

m - mean
$p(x)$ - p.d.f

if $p(x) \sim N(m, \sigma^2)$ $I = \frac{1}{\sigma^2}$

- with measurement the σ^2 decreases and information increases

Harkevich : - pragmatic level

P_{prior} - The probability of achieving a goal before receiving a message
P_{post} - Probability of ach. a goal after receiving a message

$$I_{pragm} = \log \frac{P_{post}}{P_{prior}}$$

Discrete Memoryless sources

- $X = \{ x_1, x_2, \dots, x_m \}$ $\quad Q(x_i) = \log \frac{1}{P(x_i)}$ (sh.)

- $H(X) = \sum_{i=1}^{M} P(x_i) Q(x_i) = - \sum_{i=1}^{m} P(x_i) \log P(x_i) \quad \frac{sh}{symb.}$ — Entropy

- If transmission rate of the source is $v \left(\frac{symb}{sec} \right)$ then

  $V \cdot H(X)$ — is the information flux or information rate $\left(\frac{sh}{s} \right)$

— Equivalent to Boltzmann' formula

$$\mathcal{E} \cdot k \ln \Omega$$

energy

k — Boltzman constant
Ω — number of states accessible to the system

— Properties of $H(S)$

- Continuous function in probability
- Symmetry — the order of symbols is irrelevant
- Max value — is achieved when all symbols are equally probable
- Additivity — Entropy of the union of independent events is equal to the sum of their entropies

$$H(X) - \log m = \sum_{i=1}^{m} P(x_i) \log \frac{1}{P(x_i)} - \sum_{i=1}^{m} P(x_i) \log m$$

$$= \sum_{i=1}^{m} P(x_i) \log \frac{1}{P(x_i) m}$$

$$\leq \frac{1}{\ln 2} \sum_{i=1}^{m} P(x_i) \left(\frac{1}{P(x_i) m} - 1 \right)$$

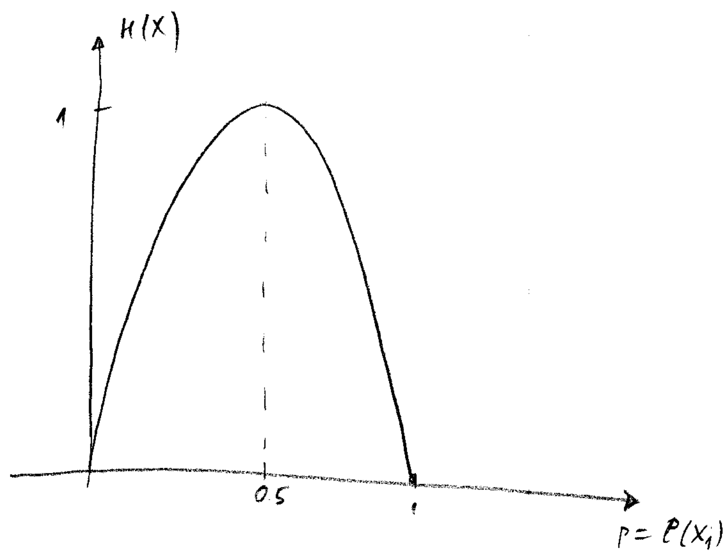
$$= \frac{1}{\ln 2} \left(\sum_{i=1}^{m} \frac{1}{m} - \sum_{i=1}^{m} P(x_i) \right)$$

$$= \frac{1}{\ln 2} (1 - 1) = 0$$

$$0 \leq H(X) \leq \log m$$

Example: $m = 2$ $X = \{x_1, x_2\}$

$$H(p) = -p \log p - (1-p) \log (1-p)$$



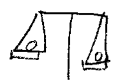
Application in solving logical problems

- Given positive integer smaller than q, what is the average number of questions required to determine the number.

  The uncertainty is: $\log q$

  - By guessing each number individually we get $H(\frac{1}{q})$ sh per question
  - By grouping in groups of we get 1 sh per question

- Finding the counterfitted coin
  - one coin has lower weight.
  - scale:



  - ternary answers $\log 3$ shannon per measurement
  - for 3^k coins k - measurements are necessary.