Information Theory (III)

Qi Zhang

Aarhus University School of Engineering

07/02/2014

AARHUS UNIVERSITET
INGENIØRHØJSKOLEN

AARHUS UNIVERSITET
INGENIØRHØJSKOLEN

Shannon Source Coding Theorem

- Source coding theorem determines a limit to possible data compression.
- Source entropy is related to the analysis of source coding theorem.
  - Assuming DMS emits a large number of symbols taken from an alphabet $A = \{x_1, x_2, \ldots, x_M\}$ in the form of a sequence of n_f symbols.
  - Priori probability of each symbol is $P(x_i),\ i = 1, \ldots, M$ and there is

$$\sum_{i}^{M} P(x_i) = 1.$$

  - A particular sequence $\mathbf{s} = s_1 s_2 \ldots s_{n_f}$ with probability

$$P(s_1 s_2 \ldots s_{n_f}) = P(s_1)P(s_2)\ldots P(s_{n_f})$$

as the symbols are statistically independent from each other.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Shannon Source Coding Theorem

- Consider a very long sequence **s**. Typically, in sequence **s** the symbol x_1 will appear $\approx n_f P(x_1)$ times, symbols x_2 will appear $\approx n_f P(x_2)$ times, ..., symbols x_M will appear $\approx n_f P(x_M)$ times.
- Hence, the probability of such **typical sequence** is roughly

$$P(\mathbf{s}) \approx P_{typ} = P(x_1)^{n_f P(x_1)} \dots P(x_M)^{n_f P(x_M)} = \prod_{i=1}^{M} [P(x_i)]^{n_f P(x_i)}$$

- It can prove that $P(\mathbf{s}) \approx 2^{-n_f H(X)}$. Hence there are $2^{n_f H(X)}$ typical sequences.
- **Typical sequences** are those with the maximum probability of being emitted by the information source.
- **Non-typical sequences** are those with very low probability of occurrence.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Shannon Source Coding Theorem

- Even though there are the total M^{n_f} possible sequences which can be emitted by information source alphabet $A = \{x_1, x_2, \ldots, x_M\}$, ONLY $2^{n_f H(X)}$ sequences have a significant probability of occurrence.
- When assuming that only $2^{n_f H(X)}$ sequences are transmitted instead of the total possible number of them, an error of magnitude ε is made. The introduced error can be arbitrary small if $n_f \to \infty$.
- It means that the source information can be transmitted using a significant lower number of sequences than the total possible number of them.
- If only $2^{n_f H(X)}$ sequences are to be transmitted and using a binary format of representation information, there will be $n_f H(X)$ bits needed for representing this information.
- So each symbol can be represented by $H(X)$ bits.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Shannon Source Coding Theorem

- For a M-ary DMS emitting equally likely symbols, there is

$$H(X) = log_2 M$$

- then

$$2^{n_f H(X)} = 2^{n_f log_2 M} = M^{n_f}$$

- In this case, the number of the typical sequences for a DMS with equally likely symbols is equal to the maximum possible number of sequences that this source can emit.
- For a DMS with independent symbols, compression of the information is possible only if the symbols of this source are not equally likely.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Source coding example

- **Example**: Let A be a 4-ary source $\{a_0, a_1, a_2, a_3\}$, we can use two binary digits to represent each source symbol. If we know that the probabilities of each symbol as follows. What is the entropy of the source?

$$P(a_0) = 0.5 \quad P(a_1) = 0.3 \quad P(a_2) = 0.15 \quad P(a_3) = 0.05$$

- **Solution**:

$$H(A) = \sum_{i=0}^{3} P(a_i) * log_2 \frac{1}{P(a_i)} = 1.6477$$

- The efficiency of the uncoded source is $H(A)/2 = 0.82385$

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Source coding example

- Instead of using 2 bits for each symbol, we can encoder the source by

$$P(a_0) = 0.5 \quad C(a_0) \rightarrow 0$$
$$P(a_1) = 0.3 \quad C(a_1) \rightarrow 10$$
$$P(a_2) = 0.15 \quad C(a_2) \rightarrow 110$$
$$P(a_3) = 0.05 \quad C(a_3) \rightarrow 111$$

  What's the average number of bits in the new coded word?

- **Solution**:

$$\overline{L} = \sum_{i=0}^{3} P(a_i)L(a_i) = .5(1) + .3(2) + .15(3) + .05(3) = 1.70$$

- The efficiency of the coded source is $H(A)/\overline{L} = 0.96924$

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Source coding summary

- For a DMS emitting an alphabet $A = \{x_1,\ x_2, \ldots, x_M\}$
  - The arbitrary information sources can have a considerable range of possible entropies. $0 \leq H(X) \leq log_2(M)$;
  - The entropy of a source is the average information carried per symbol;
  - As there is a cost to transmit or store each symbol, it is desirable to use minimum number of symbols to represent the information.
  - It is possible to compress the information provided the source only if the symbols of this source are not equally likely, i.e., $H(X) < log_2(M)$.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Prefix codes and instantaneous Decoding

- Let's look at this sequence of letters:
  - IFIWANTEDTOPICKONE
  - IF I WANTED TO PICK ONE vs. IF I WANT ED TO PICK ONE
- The English language is not generally self-punctuating.
- **Prefix code** is a code that has the property of being self-punctuating.
  - It has punctuation built into the structure.
  - It is accomplished by designing the code such that no codeword is a prefix of another (longer) codeword.
  - It is instantaneously decodable.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

- **Example**: Let the encoded map pairs of symbols into the codewords shown below. Please decode the sequence: 10000011111111011101, assuming the codewords are transmitted bit serially from left to right.

$\langle x_i, x_j \rangle$	$P(x_i, x_j)$	b_m	$\langle x_i, x_j \rangle$	$P(x_i, x_j)$	b_m
x_1x_1	.25	00	x_3x_1	.075	1101
x_1x_2	.15	100	x_3x_2	.045	0111
x_1x_3	.075	1100	x_3x_3	.0225	111110
x_1x_4	.025	11100	x_3x_4	.0075	1111110
x_2x_1	.15	101	x_4x_1	.025	11101
x_2x_2	.09	010	x_4x_2	.015	111101
x_2x_3	.045	0110	x_4x_3	.0075	11111110
x_2x_4	.015	111100	x_4x_4	.0025	11111111

- **Solution**:
  - $100, 00, 0111, 111110, 11101$;
  - which decodes as $x_1x_2x_1x_1x_3x_2x_3x_3x_4x_1$.

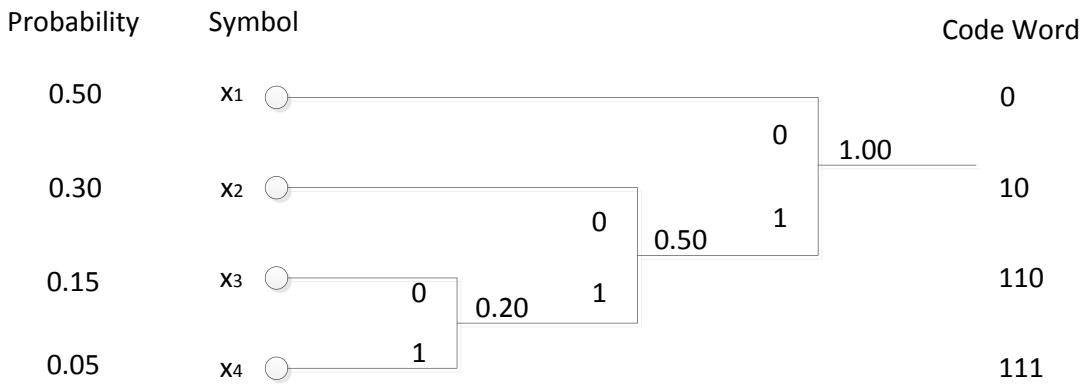
AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Construct a Huffman code example

- **Example**: Construct a Huffman code for the 4-ary source alphabet x_1, x_2, x_3, x_4 with probability

$$P(x_1) = 0.5 \quad P(x_2) = 0.3 \quad P(x_3) = 0.15 \quad P(x_4) = 0.05$$

- The constructed Huffman tree:
  - Start from small probabilities
  - Form a tree
  - Assign 0 to higher branch, 1 to lower branch



AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Huffman coding

- Huffman codes are lossless data compression codes;
- Huffman codes are widely used in data communications, speech coding, and video or graphical image compression;
- Huffman codes can deliver codeword sequences which asymptotically approach the source entropy.
- Huffman codes generally have variable length codewords.
- Huffman codes belong to prefix codes.

AARHUS UNIVERSITET
INGENIØRHØJSKOLEN

Preview of Channel coding theorem

- Taking BSC channel as example here. The input **X** and the output **Y** are the sequences of n symbols.
- With error probability p of BSC, the output **Y** will differ in np positions comparing with **X**.
- The number of sequence of n symbols with difference in np position is $\binom{n}{np} \approx 2^{n\Omega(p)}$.
- This result indicates that for each input sequence of n symbols, there exists $2^{n\Omega(p)}$ possible output sequences as a result of the errors introduced by the channel.
- The total number of the typical output sequences of this channel $2^{nH(Y)}$.
- Use only uniquely distinguishable input sequences, i.e., divide the set of $2^{nH(Y)}$ sequences into disjoint sets of size $2^{n\Omega(p)}$, the maximum number of such sets is

$$\frac{2^{nH(Y)}}{2^{n\Omega(p)}} = 2^{n[H(Y)-\Omega(p)]} = 2^{nI(X;Y)}$$

- This derivation outlines an upper bound on the capacity.

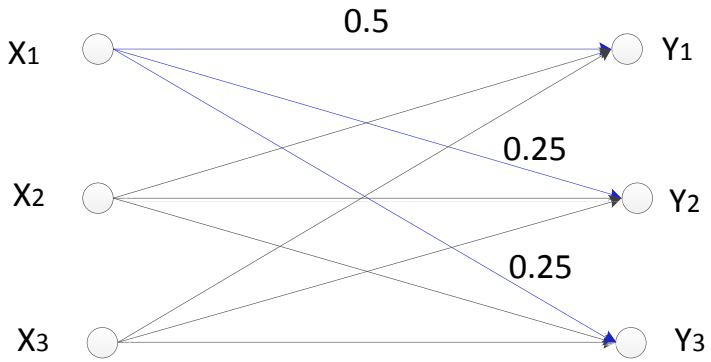
AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Channel coding theorem

- **Channel Coding Theorem**: If the transmission rate R fits the condition $R < C$, then for an arbitrary value $\varepsilon > 0$, there exists a code with block length n that makes the error probability of the transmission be less than ε. If $R > C$ then there is no guarantee of reliable transmission; namely there is no guarantee that the arbitrary value of ε is a bound for the error probability.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Example 1.11: Calculate Channel capacity



- **Example 1.11**: Determine the channel capacity of the channel whose model is shown above. Assuming the input symbols are equally likely, and the channel transition probabilities are listed blow.

$$\begin{array}{lll} P(y_1/x_1) = & P(y_2/x_2) = & P(y_3/x_3) = 0.5 \\ P(y_1/x_2) = & P(y_1/x_3) = & 0.25 \\ P(y_2/x_1) = & P(y_2/x_3) = & 0.25 \\ P(y_3/x_1) = & P(y_3/x_2) = & 0.25 \end{array}$$

Calculate the channel capacity.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Shannon's Channel Capacity Theorem

- Channel Capacity in bits per second of additive Gaussian noise (AWGN) channel:

$$C = B log_2 \left(1 + \frac{S}{N_0 B}\right)$$

  - where B is the bandwidth, S is the signal power, N_0 is the power density of the noise;

- The channel capacity, C increases as the available bandwidth B increases and as the signal to noise ratio $\frac{S}{N}$ increases (improves).
- In a single formula it highlights the interplay between 3 key system parameters:
  - channel bandwidth
  - average received signal power
  - noise power at the channel output

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Shannon Limit

- Channel Capacity in bits per second of additive Gaussian noise (AWGN) channel:

$$C = Blog_2\left(1 + \frac{S}{N_0 B}\right)$$

  - where B is the bandwidth, S is the signal power, N_0 is the power density of the noise;
  - Rewrite the above equation using E_b average energy per bit:

$$\frac{C}{B} = log_2\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right)$$

  - There is $\lim_{x\to 0}(1+x)^{1/x} = e$;
  - Based on $log_2(1+x) = x\frac{1}{x}log_2(1+x) = xlog_2\left[(1+x)^{1/x}\right]$,

$$\frac{C}{B} = \frac{E_b}{N_0}\frac{C}{B}log_2\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right)^{N_0 B/CE_b}$$

$$\Rightarrow \quad 1 = \frac{E_b}{N_0}log_2\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right)^{N_0 B/CE_b}$$

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

- If $\frac{C}{B} \to 0$,

$$\frac{E_b}{N_0} = \frac{1}{log_2(e)} = 0.6931$$

$$\left(\frac{E_b}{N_0}\right)_{dB} = 10log_{10}(0.6931) = -1.59\ dB$$

- -1.59dB is Shannon limit.
- It means that if the ratio E_b/N_0 is kept sightly higher than -1.59dB, it is possible to have error-free transmission by sophisticated coding technique.
- In other words, if the ratio E_b/N_0 is below -1.59dB. It is impossible to have error-free transmission, no matter how sophisticated coding is used.

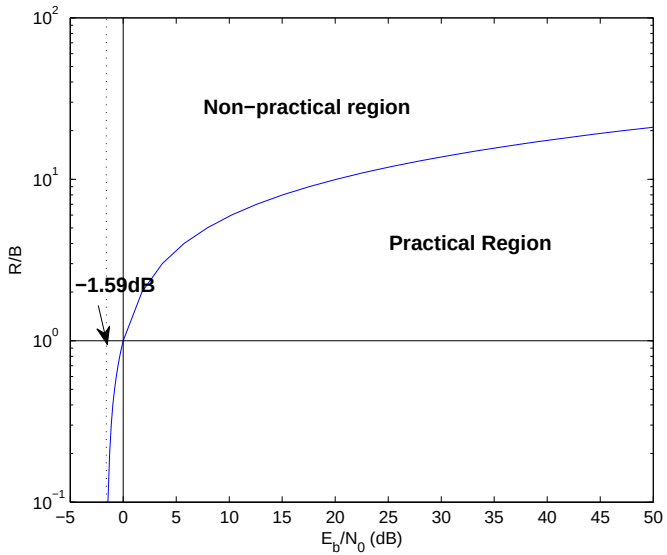
AARHUS UNIVERSITET INGENIØRHØJSKOLEN

$$C = Blog_2\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right)$$

Rewrite the equation into:

$$2^{C/B} = 1 + \frac{E_b}{N_0}\left(\frac{C}{B}\right)$$

Let $R = C$,



- The curve shows when $R/B \to 0$, E_b/N_0 reaches Shannon limit.
- For each value of R/B, there exists a corresponding bound which can be read from the curve.

AARHUS UNIVERSITET INGENIØRHØJSKOLEN

Summary of Information Theory

- Information, I_i.
- Entroy, $H(X)$.
- Information channels, e.g., BSC, BEC. And their corresponding channel model.
- Different probabilities, i.e., $p(x_i)$, $p(y_j)$, $p(x_i, y_j)$, $p(y_j/x_i)$, $p(x_i/y_j)$.
- Different Entropies, i.e., $H(X)$, $H(Y)$, $H(X, Y)$, $H(Y/X)$, $H(X/Y)$.
- Mutual information $I(X; Y)$ and its properties.
- Shannon source coding theorem.
- Shannon channel coding theorem
- Shannon limit.

AARHUS UNIVERSITET
INGENIØRHØJSKOLEN