Subject ELEC3203 No.

Digital Coding & Transmission

Klein

If you find, please e-mail: lk1n16@ecs.soton.ac.uk

Basic+note

Motivate higher office efficiency!

A5

8mmx21Lines

deli得力

80页 无线装订本

No.7654

A5 | 规格：210x148mm

得力集团有限公司

DELI GROUP CO., LTD.

地址：浙江宁海得力工业园

Add:Deli Industry Zone, Ninghai County, Zhejiang Province,China

电话(Tel): 0574-65278888

传真(Fax): 0574-65273660

Http://www.nbdeli.com

MADE IN CHINA

执行标准：QB/T1438 合格

6 921734 976543

2016/10/03rd

Information Theory

- Measure of information
- Information capacity of channel
- Coding as a means of utilising channel capacity

to communicate information between geographically separated locations

Aim of telecommunications: via a communication channel of adequate quality (at certain rate reliab

The transmission is based on digital data, which is something obtained from analogue quantities by:

① Sampling (Nyquist: sampling with at least twice the maximum frequency)

② Quantisation

§ General Transmission Scheme

input → Source Encoding → Channel Encoding → Modulation → Channel

CODEC 编码解码器　　MODEM 调制解调器

Channel → Demodulation → Channel Decoding → Source Decoding → output

§ What is information

General digital source is characterised by:

① Source alphabet (message or symbol set): $m_1, m_2, \cdots, m_q$ q: size

② Probability of occurrence (symbol probabilities): $p_1, p_2, \cdots, p_q$

③ Symbol rate (symbols/s or Hz)

④ Probabilistic interdependence of symbols (correlation of symbols)
- memory ?
- memoryless ?

$\lim_{p_i \to 1} I(m_i) \to 0$

Amount of information: $I(m_i) = \log_2 \frac{1}{\cancel{q_i}\; p_i} = -\log_2 \cancel{q_i}\; p_i$ [bits]

$\lim_{p_i \to 0} I(m_i) \to 1$ Never occur

No.

Date

Amount of information: $I(m_i) = \log_2 \frac{1}{P_i} = -\log_2 P_i$ [bits]

※ Message that convey information that is certain to happen and already known by recipient contain NO REAL INFORMATION.

Eg. A certain message of probability 1: $I(m) = \log_2 1 = 0$

Information content of an event is connected with uncertainty or inverse of probability. The more unexpected (smaller probability) the event is, the more information it contains

§ Shannon limit

Shannon's theorem: If the rate of information from a source does not exceed the capacity of a communication channel, then there exists a coding technique such that the information can be transmitted over the channel with arbitrarily small probability of error, despite the presence of noise.

§ Source and Source Coding

→ map each $S(k)$ to a code word or bit pattern
convert $\{S(k)\}$ [symbols/s] into $\{b_i\}$ [bits/s]

※ $\boxed{\text{digital source}} \xrightarrow[\text{symbols/s}]{\{S(k)\}} \boxed{\text{source coding}} \xrightarrow[\text{bits/s}]{\{b_i\}}$

Sequence of symbols $\{S(k)\}$ ← time index

Characters:

① Symbol set or source alphabet: $S = \{m_i, 1 \le i \le q\}$

② Probability of occuring for m_i: P_i, $1 \le i \le q$

③ Symbol rate R_S [symbols/s]

④ Source memory or ~~trade~~ interdependency in $\{S(k)\}$

$\{S(k)\} = m_j$ (probability) $\{S(k-1)\} = m_i$

Source coding is concerned with coding each $\boxed{m_i \in S}$ with a unique code word or bit pattern

---

No.

Date

§ Physical Interpretation

Information content of a symbol or message is equal to minimum number of binary digits required to encode it and, hence, has a unit of bits.

q $\boxed{\text{equiprobable}}$ symbols m_i, $1 \le i \le q$ → minimum number of bits: $\log_2 q$ ← information content

Use $\log_2 q$ bits for each symbol is called $\boxed{\text{Binary Code Decimal}}$

§ Information of Memoryless Source

$S(k)$ is independent of the previous symbols $s(k-i)$, $i \ge 1$

※※ N是长度！q是symbol的个数！ $q \le N$

Assume $\{S(k)\}_{k=1}^{N}$ ← sequence of length, symbol m_i appears $P_i \cdot N$ times in sequence

→ Information contribution from ith symbol m_i: $I_i = P_i \cdot N \cdot \log_2 \frac{1}{P_i}$

→ Total information of symbol sequence (of length N): $I_{total} = \sum_{i=1}^{q} I_i = \sum_{i=1}^{q} P_i \cdot N \cdot \log_2 \frac{1}{P_i}$

→ Average information per symbol (熵 entropy): $\frac{I_{total}}{N} = \sum_{i=1}^{q} P_i \cdot \log_2 \frac{1}{P_i}$ (bits/symbol)

§ Entropy and Information Rate

Memoryless source entropy: $H = \sum_{i=1}^{q} P_i \cdot \log_2 \frac{1}{P_i}$ [bits/symbol]

Information rate: $R = R_s \cdot H$ [bits/sec] (R_s ← symbol rate)

Entropy is a fundamental physical quantity of the source, quantifies average information conveyed per symbol.

Information rate is a fundamental physical quantity of the source, tells you how many bits/s information the $\boxed{\text{source really needs}}$ to send out.

necessary → $R \le R_b$ ← rate you send (appear waste)

Information rate R is always smaller ~~than~~ or equal to the average output bit rate R_b

$$R_b = R_s \cdot \log_2 q \quad \text{[Binary Code Decimal]}$$

For memoryless source, binary coded decimal source coding is generally "waste"

Code Symbols to Achieve Efficiency

☆ $R_b = R$? X $\begin{cases} \text{For equiprobable } R_b = R_s \cdot \log_2 q \\ \text{Otherwise: } R_b = \sum \text{length} * P_{\text{length}} \end{cases}$

长度 ; 这个长度出现的概率

§ Maximum Entropy for q-ary Source

$H = -\sum_{i=1}^{q} P_i \cdot \log_2 P_i \rightarrow$ Lagrangian : $\mathcal{L} = (\sum_{i=1}^{q} -P_i \log_2 P_i) + \lambda \cdot (1 - \sum_{i=1}^{q} P_i)$

Yields : $\frac{\partial \mathcal{L}}{\partial P_i} = -\log_2 P_i - \log_2 e - \lambda = 0$

Since $\begin{cases} \log_2 P_i = -(\log_2 e + \lambda) \text{ is independent of } i \\ \sum_{i=1}^{q} P_i = 1 \end{cases}$

Entropy of a q-ary source is maximised for equiprobable symbols with $P_i = \frac{1}{q}$

§ Coding Efficiency

$R = R_s \cdot H \le R_s \cdot \log_2 q$ $\boxed{\begin{array}{l} R = R_s \cdot \log_2 q \\ \text{Only for equiprobable symbols} \end{array}}$

Source entropy is bound by maximum entropy as $0 \le H \le \log_2 q$

☆ BCD only achieves most efficient source coding for equiprobable symbols.

Coding efficiency : code efficiency = $\frac{\text{source information rate } R}{\text{average source output rate } R_b}$

Shannon's source coding theorem : with an efficient source coding, a coding efficiency of almost 100% can be achieved.

For source with equiprobable symbols :

$P_i = \frac{1}{q}$, $1 \le i \le q$ Source entropy is maximised : $H = \log_2 q$ bits/symbol

$CE = \frac{R}{R_b} = \frac{H}{\log_2 q} = \frac{\log_2 q}{\log_2 q} = 100\%$

For source with non-equiprobable symbols :

Coding each symbol ~~with~~ into $\log_2 q$-bits codeword is not efficient, as $H < \log_2 q$

$CE = \frac{H}{\log_2 q} < 100\%$

☆ Assign number of bits to a symbol according to its information content [variable-bits codewords]

§ SHANNON - FANO Coding

① Order symbols m_i in descending order of probability

② Divide symbols into subgroups such that the subgroup's probabilities are as close as possible

③ Allocating codewords : sign bit 0 to tops subgroup and bit 1 to bottom subgroup

④ Iterate step 2 and 3 as long as there is more than one symbol in any subgroup

⑤ Extract variable-length codewords from the resulting tree (top-down)

Note: No codewords forms a prefix for any other codeword, so they can be decoded unambiguously

[Example] ☆ less probable symbols are coded by longer code words while high probable symbols are assigned short codes

approx length	I_i (bits)	Symb m_i	Prob P_i	Coding Steps 1	2	3	4	Codeword
2	1.89	m_1	0.27 (0.47)	0	0			00
2	2.32	m_2	0.20	0	1			01
3	2.56	m_3	0.17 (0.33)	1	0	0		100
3	2.64	m_4	0.16	1	0	1		101
4	4.06	m_5	0.06 (0.2)	1	1	0	0	1100
4	4.06	m_6	0.06	1	1	0	1	1101
4/5	4.64	m_7	0.04	1	1	1	0	1110
4/5	4.64	m_8	0.04	1	1	1	1	1111

Entropy : $H = \sum_{i=1}^{q} P_i \cdot \log_2 \frac{1}{P_i} = 2.6906$ [bits/symbol]

Average codeword length : $0.47 \cdot 2 + 0.33 \cdot 3 + 0.2 \cdot 4 = 2.73$ [bits/symbol]

$CE = \frac{\text{source information rate } R}{\text{average source output rate } R_b} = \frac{R_s \cdot H}{R_s \cdot 2.73} = \frac{2.6906}{2.73} = 98.56\%$

[Comparison]

With 3-bits equal length codewords : ~~$\log_2 8$~~ $\sum_{i=1}^{q} P_i \cdot 3 = 3$

$CE = \frac{2.6906}{3} \times 100\% = 89.69$

§ Huffman Coding

① Arrange symbols in descending order of probabilities.

② Merge the two least probable symbols (or subgroups) into one subgroup

③ Assign '0' and '1' to the higher and less probable branches in the sub respectively

④ If there is more than one symbol or sub left, return to step 2

⑤ Extract the Huffman code words from the different branches (bottom-up)

[Example]

m7 0, m8 1 | m1, m2, m3, m4, m78, m5 0, m6 1 | m1, m2, m3, m4, m56 0, m78 1 | m1, m2, m5678, m3 0, m4 1 | m34, m1, m2 0, m5678 1 | m25678, m34 0, m1 1 | m134 0, m25678 1

Symb m_i	Prob P_i	Coding steps 1	2	3	4	5	6	7	reverse
m_1	0.27						1	0	01
m_2	0.20					0		1	10
m_3	0.17				0		0	0	000
m_4	0.16				1		0	0	001
m_5	0.06		0	0		1		1	1100
m_6	0.06		1	0		1		1	1101
m_7	0.04	0		1		1		1	1110
m_8	0.04	1		1		1		1	1111

No codeword forms a prefix for any other codeword.

$H = 2.690\,6$

$CE = \frac{H}{2.73} = 98.56\%$

☆ 编码倒置

§ Summary

digital Source → {S(k)} symbols/s → source coding → {bit} bits/s

① Entropy coding {Shannon-Fano, Huffman} ← Both decode unambiguously; which assigns number of bits to a symbol as close as possible to its information contents

② Both Sxx and Bxx can code memory-less ~~sequence~~ sources such that the emitted bit sequence carries Maximum of information

③ For $\{S(k)\}$, $1 \le k \le q$ (size), the longest codeword could be up to $q-1$ bits. ☆ Prohibitive for large alphabets.

④ Huffman coding gives a different codeword assignment, the CE however is nearly idental.

Data rate → information rate → efficient encoding → data rate in minimised
As close to information rate as possible

Constraints: { Number of bits can only be integer
No code word forms a prefix for any other code word

Sources with Memory

Most real world sources exhibit memory, resulting in coorelated source signals. This property is retained during sampling and quantisation.

↳ This implies that the signal exhibits some form of ~~redundany~~ redundancy, which should be exploited when the signal is coded

§ Model Source Memory

Memory can be completely modelled by a stochastic (随机的) probablistic Markov process

- Source with memory that emits a sequence of symbols $\{S(k)\}$ with time index k.
- First order Markov process:
  The current symbol depends only on the previous symbol. $p(S(k) \mid S(k-1))$
- N-th order Markov process:
  The current symbol depends on N previous symbols, $p(S(k) \mid S(k-1), S(k-2), \dots, S(k-N))$

Alternatively, if $S(k)$ is influenced by $S(k-1)$ up to $S(k-N)$, then it may be modelled by predictive model:

contains information of S(k) that can be predicted by S(k)…S(k−N)

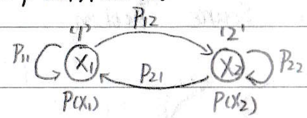
$$S(k) = \underbrace{f(S(k-1), \dots, S(k-N))}_{\text{contains information of } S(k) \text{ that can be predicted}} + \varepsilon(k)$$ ← contains new information of S(k)

§ Two-state First Order Markov Process

Source $S(k)$ can only generate two symbols, $X_1 = 1$ and $X_2 = 2$, their probability explicitly depends on the previous state (i.e. $p(S(k) \mid S(k-1))$)

Transition probility matrix:

$$T = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$



$P_{12} : P_{1 \to 2}$

Probilities of occurrence (prior probilities) for states X_1 and X_2: $\begin{cases} P_1 = P(X_1) \\ P_2 = P(X_2) \end{cases}$

i.e. $\begin{cases} p(S(0) = 1) = P(X_1) \\ p(S(0) = 2) = P(X_2) \end{cases}$

Transition Probilities: from state X_1 are given by the conditional probabilities

$\begin{cases} P_{12} = P(X_2 \mid X_1) \\ P_{11} = P(X_1 \mid X_1) = 1 - P(X_2 \mid X_1) \end{cases}$

i.e. $p(S(k) = j \mid S(k-1) = i) = P_{ij}$

Entropy H_i for state X_i, $i=1,2$: ✓ average information carried by the symbols emitted in state X_i

$$H_i = -\sum_{j=1}^{2} P_{ij} \cdot \log_2 P_{ij} = -P_{i1} \cdot \log_2 P_{i1} - P_{i2} \cdot \log_2 P_{i2} \quad (\text{bits/symbol})$$

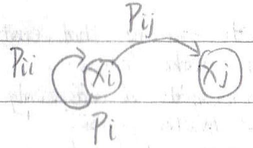
Overall entropy H:

$$H = \sum_{i=1}^{2} P_i H_i = -\sum_{i=1}^{2} P_i \sum_{j=1}^{2} P_{ij} \cdot \log_2 P_{ij} \quad [\text{bits/symbol}]$$

Entropy decreases as correlation increases

Information rate : ~~R_b~~ $R = R_s \cdot H$ (bits/second)   (R_s: symbols/second)

§ Entropy for q-state 1st order Markov Source



Markov source with q symbols $X_i = i$, $1 \le i \le q$

P_{ij} is the transition probability from X_i to X_j

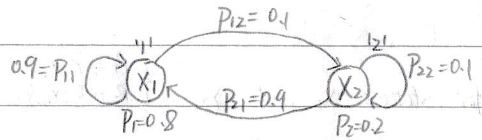
Symbol entropy : $H_i = -\sum_{j=1}^{q} P_{ij} \cdot \log_2 P_{ij}$ [bits/symbol]

Source entropy : $H = -\sum_{i=1}^{q} P_i \sum_{j=1}^{q} P_{ij} \cdot \log_2 P_{ij} = +\sum_{i=1}^{q} P_i \cdot H_i$ [bits/symbol]

↑ probabilities of occurrance (prior probability)

Average source information rate : $R = R_s \cdot H$ [bits/second]

§ A 2-state 1st Order Markov Source Problem



$$H^m = 0.8 \cdot \log_2 \frac{1}{0.8} + 0.2 \cdot \log_2 \frac{1}{0.2} = 0.7219$$

[Source Entropy]

$$H = \sum_{i=1}^{q} P_i \cdot H_i = \sum_{i=1}^{q} P_i \cdot \sum_{j=1}^{q} P_{ij} \log_2 \frac{1}{P_{ij}} = P_1 \cdot \sum_{j=1}^{q} P_{1j} \log_2 \frac{1}{P_{1j}} + P_2 \cdot \sum_{j=1}^{q} P_{2j} \log_2 \frac{1}{P_{2j}}$$

$$= 0.8 \cdot \left(0.9 \cdot \log_2 \frac{1}{0.9} + 0.1 \cdot \log_2 \frac{1}{0.1}\right) + 0.2 \cdot \left(0.9 \cdot \log_2 \frac{1}{0.9} + 0.1 \cdot \log_2 \frac{1}{0.1}\right)$$

$$= 0.469 \ [\text{bits/symbol}]$$

[Average information] For 1-symbol, 2-symbol, 3-symbol sequence

① For 1-symbol sequence "1" or "2"

$$-P(\text{"1"}) \log_2 P(\text{"1"}) - P(\text{"2"}) \log_2 P(\text{"2"}) = -0.8 \cdot \log_2 0.8 - 0.2 \cdot \log_2 0.2 = 0.7219 \ [\text{bits/symbol}]$$

② For 2-symbol sequence "11", "12", "21", "22"

$P(\text{"11"}) = P(\text{"1"}) \times P_{11} = 0.8 \times 0.9 = 0.72$

$P(\text{"12"}) = P(\text{"1"}) \times P_{12} = 0.8 \times 0.1 = 0.08$

$P(\text{"21"}) = P(\text{"2"}) \times P_{21} = 0.2 \times 0.9 = 0.18$

$P(\text{"22"}) = P(\text{"2"}) \times P_{22} = 0.2 \times 0.1 = 0.02$

$$-0.72 \log_2 0.72 - 0.08 \log_2 0.08 - 0.18 \log_2 0.18 - 0.02 \log_2 0.02 = 1.191 \ [\text{bits}]$$

$$H^{(2)} = \frac{1.191}{2} = 0.5955 \ [\text{bits/symbol}]$$

③ For 3-symbol sequence 111, 112, 121, 122, 211, 212, 221, 222

$P(\text{"111"}) = P(\text{"1"}) \times P_{11} \times P_{11} = 0.8 \times 0.9 \times 0.9 = 0.648$

$P(\text{"112"}) = P(\text{"1"}) \times P_{11} \times P_{12} = 0.8 \times 0.9 \times 0.1 = 0.072$

$P(\text{"121"}) = P(\text{"1"}) \times P_{12} \times P_{21} = 0.8 \times 0.1 \times 0.9 = 0.072$

$P(\text{"122"}) = P(\text{"1"}) \times P_{12} \times P_{22} = 0.8 \times 0.1 \times 0.1 = 0.008$

$P(\text{"211"}) = P(\text{"2"}) \times P_{21} \times P_{11} = 0.2 \times 0.9 \times 0.9 = 0.162$

$P(\text{"212"}) = P(\text{"2"}) \times P_{21} \times P_{12} = 0.2 \times 0.9 \times 0.1 = 0.018$

$P(\text{"221"}) = P(\text{"2"}) \times P_{22} \times P_{21} = 0.2 \times 0.1 \times 0.9 = 0.018$

$P(\text{"222"}) = P(\text{"2"}) \times P_{22} \times P_{22} = 0.2 \times 0.1 \times 0.1 = 0.002$

$$-0.648 \log_2 0.648 - 0.072 \log_2 0.072 - \cdots - 0.002 \log_2 0.002 = 1.0079 + 0.6520 = 1.66$$

$$H^{(3)} = \frac{1.66}{3} = 0.5533 \ [\text{bits/symbol}]$$

§ Compare Memory and Memoryless Sources

Same : ① symbol set ② probability of occurring ③ symbol rate

$$H^{(ml)} \ge H^{(m)} \quad \& \quad R^{(ml)} \ge R^{(m)}$$

m: memory
ml: memoryless

Directly code memory source $\{S(k)\}$ by entropy coding you ONLY Get "1-symbol-sequence entropy" $H^{(1)} = H^{(ml)}$

$$R_b = R_s \cdot H^{(1)} \gg R_s \cdot H^{(m)}$$

You sent at rate far far larger than the source information rate $R^{(m)}$

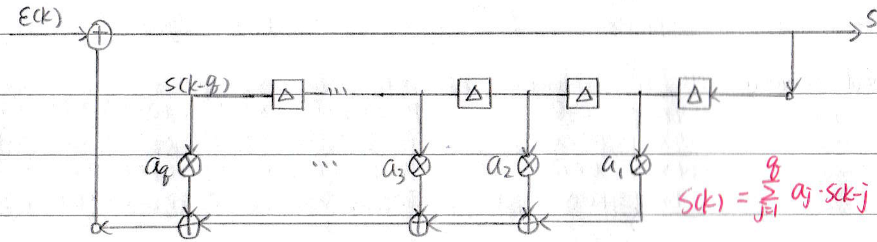
§ Predictive Models

An qth order predictive model with parameter vector a:

$$\hat{s}(k) = E[s(k) \mid s(k-1), s(k-2), \dots, s(k-q)] + \varepsilon(k)$$

$$= f(s(k-1), s(k-2), \dots, s(k-q); a) + \varepsilon(k)$$

E.g. qth order linear autoregressive (AR) model: (自回归)



$$s(k) = \sum_{j=1}^{q} a_j \cdot s(k-j) + \varepsilon(k)$$

- Aim to get residual (剩余) sequence $\{\varepsilon(k)\}$ uncorrelated and zero-mean
- This parametric model is widely used, e.g. in speech (source) coding — transmit a_j and $\varepsilon(k)$ instead of $s(k)$

[例] 一离散信号由 0,1,2,3 四个符号组成，它们的概率分别为 3/8，1/4，1/4，1/8，且每个符号的出现都是独立的。试求某信息 20102 ..... 20210 的信息量（"0"出现23次，"1"出现14次，"2"出现13次，"3"出现7次，共有57个符号

信息量：$I = 23\log_2\frac{8}{3} + 14\log_2 4 + 13\log_2 4 + 7\log_2 8 = 108$ [b] ← 存在误差，误差随消息序列中符号数的增加而减少

算术平均信息量：$\bar{I} = I/$符号数 $= 108/57 = 1.89$ [b/symbol]

用熵的概念来计算 $H = -\frac{3}{8}\log_2\frac{3}{8} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} = 1.906$ [b/symbol] $\Rightarrow I = H \cdot N = 108.64$ [b] (no. of symbols 57)

信号能量 $E = \int_{-\infty}^{+\infty} s^2(t)\,dt$ 若 $0 < E < \infty$，则称此信号为#能量信号

信号平均功率 $P = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} s^2(t)\,dt$

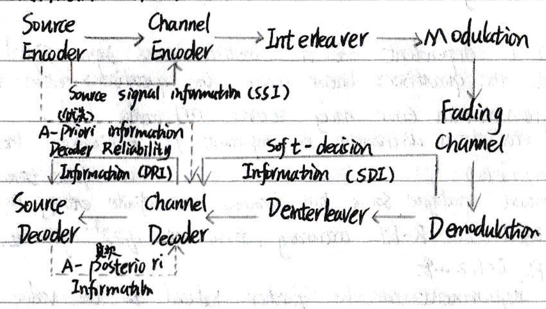
能量信号的平均功率为零 → 能量有限，时间趋于无穷

#功率信号：信号平均功率是一个有限的正值（现实中存在）

频谱密度 (frequency spectrum density) ← 能量信号 $s(t)$ 的傅里叶变换 $S(f) = \int_{-\infty}^{+\infty} s(t)\, e^{-j2\pi f t}\,dt$ $\quad Sa(\pi f \tau) = \frac{\sin(\pi f \tau)}{\pi f \cdot \tau}$

2016/10/10th

Shannonian Transceiver Schematic



Speech Signals and Speech Coding

§ Basic Characterisation of Speech Signals

Random signals such as speech, music, video, etc information signals cannot be describa by the help of analytical formule. They are typically characterised by the help of statistical characteristic such as PSD, ACF, CDF and PDF

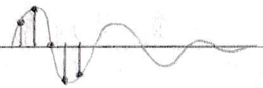
功率谱密度 Power Spectral Density; Auto correlation Function

§ Waveform Coding

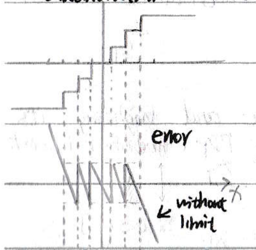
anti-aliasing low-pass filter

Digitisation of analogue speech signals

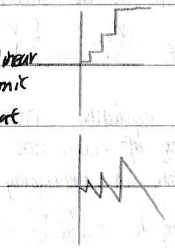
Speech in → LPF B → Sampling 2B → Quantisation → Parallel To Serial Converter → Binary Speech Bits



§ Quantisation Characteristic

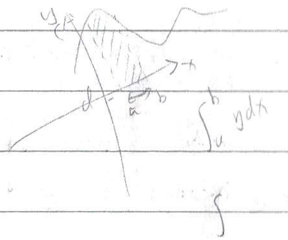


Linear quantisers exhibit a linear transfer function within the dynamic range and saturation above that



Mid-tread quantiser

Non-unifor quantiser



σ_x^2 "power" in $x[n] = E\{x^2[n]\}$ $\sigma \in$ variance
$\sigma_{e_q}^2$ "power" in $e_q[n] = E\{e_q^2[n]\}$

No.

Date

§ Quantisation Noise and Rate-Distortion Theory

The instantaneous quantisation error $e(x)$ is dependent on its instantaneous input signal level.
If the input signal's dynamic range exceeds the quantiser's linear range, the quantiser's output voltage saturates at its maximum level and the quantisation error may becom arbitarily high.
Hence the knowledge of the input signal's statistical distribution is important for minimising the overall granular and overload distortion. $\hat{x}(t) = x(t) + e(t)$
If no amplitude discreation is used, a sampled analogue soure has formally an infinite entropy, requiring an infinite transmission rate
If the analogue speech samples are quantised to R-bit accuracy, there are $q = 2^R$ legitimate samples, each of which has a probability of occurance P_i, $i = 1, 2 \ldots q$.
The above R bit/symbol channel capacity requirement can be further reduced to the value of the source's entropy given by $H(x) = -\sum_{i=1}^{q} P_i \cdot \log_2 P_i$ using entropy coding without inflicting any further coding impairment, if an infinite delay coding is acceptable, although this is not practical.

[Rate-distortion Theorem]

The so-called rate-distortion theorem, which quantifies the minimum required average bit rate R_D in terms of [bit/sample] in order to represent a rv with less than D distortion.
Explicitly, for a rv (random variable) with variance of σ_x^2 and quantised value $\hat{x}$ the distortion is defined as the mean square error (mse):

expected value

$$D = E\{(x - \hat{x})^2\} = E\{e(t)\}$$

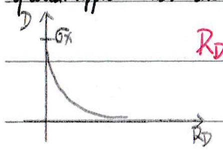
If $R_D = 0$ bits are used to quantise the quanity x, then the distortion is given by the signal's variance $D = \sigma_x^2$.

*Signal to noise ratio compares the level of a desired signal to the level of backgraund noise.
If SNR > 1, it indicates more signal than noise.

$$SNR = \frac{\text{Power of signal}}{\text{Power of noise}} = \frac{P_{signal}}{P_{noise}} = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}$$

R_D每多一个bit，distortion就会缩小一半，针对需要的D，
可以求出最小的位数.

If $R_D > 0$, then one additional bit is needed every time we want to halve the rms value of 'D', or quadrupple (4倍) the SNR (σ_x^2/D), which suggests a logarithmic relation between R_D and D.



$$R_D = \frac{1}{2}\log_2 \frac{\sigma_x^2}{D} \Rightarrow \begin{cases} R_D = \frac{1}{2}\log_2 \frac{\sigma_x^2}{D} & D \le \sigma_x^2 \\ R_D = 0 & D > \sigma_x^2 \end{cases}$$

Distortion 大于信号，没有采样的必要
$\Leftarrow$ 信噪比 即 SNR < 1

quantisation interval

$e(t)$ is uniformly distributed in the interval $[-\frac{q}{2}, \frac{q}{2}]$ if no quantiser overload is incurred

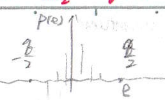
Quantisation interval: $q = \frac{2V}{2^b}$ $\begin{cases} V: \text{quantiser's linear range} \\ b: \text{number of bits quantised} \end{cases}$

σ_e^2 can be computed by squaring instantaneous error magnitude e and wighting its contribution with its probility of occurance expressed by the help of its PDF (Probability Distribution Fun) $p(e) = 1/q$ and finially integrating or averaging it over the range of $[-\frac{q}{2}, \frac{q}{2}]$

$$\sigma_e^2 = \int_{-\frac{q}{2}}^{\frac{q}{2}} e^2 p(e)\, de = \int_{-\frac{q}{2}}^{\frac{q}{2}} e^2 \frac{1}{q}\, de = \left[\frac{e^3}{3}\right]_{-\frac{q}{2}}^{\frac{q}{2}} = \frac{q^2}{12}$$

[Rms quantiser noise $\frac{q}{\sqrt{12}} \approx 0.3q$]

Correspond to



SNR was derived for uniform signal PDF & Quatization noise PDF
PDF is not uniform but known
SNR is independent of the PDF

No.

Date

which gives the noise variance

In case of uniform quantisers we can substitute $q = \frac{2V}{2^b}$ in $\sigma_q^2 = \frac{q^2}{12} = \frac{1}{12}\left(\frac{2V^2}{2^{b}}\right) = \frac{1}{3} \cdot \frac{V^2}{2^{2b}}$

Assuming a uniform signal PDF the signal's variance becomes:

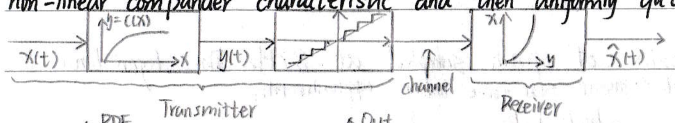
$$\sigma_x^2 = \int_{-\infty}^{+\infty} x^2 p(x)\, dx = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{2V}\, dx = \frac{1}{2V}\left[\frac{x^3}{3}\right]_{-V}^{V} = \frac{2V^3}{6E} = \frac{E^2}{3}$$

Then the SNR can be computed as: $\begin{cases} SNR = \frac{\sigma_x^2}{\sigma_q^2} = \frac{V^2}{3} \cdot \frac{2^{2R}}{V^2} \cdot 3 = 2^{2R} \\ SNR_{dB} = 10 \cdot \log_{10} 2^{2R} = 20R \cdot \log_{10} 2 = 6.02\, R\ [dB] \end{cases}$

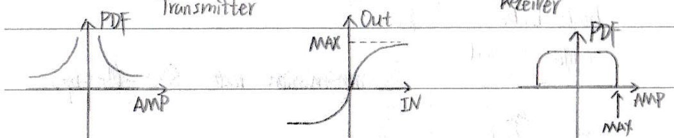
In pratice, the speech PDF is highly nonuniform and the quantiser's dynamic range can't be fully exploited, hence the equation over-estimates the expected SNR.

§ Non-uniform Quantisation: Companding

If the input signal's PDF is known and can be considered stationary, higher SNR can be achieved by approriately matched non-uniform quantisation (NUQ) than in case of uniform quantiers.
The quantisation intervals are more dense near the origin in order to quantize the typically high-probaility low-~~mang~~ magnitude samples more accurately. In contrast, the lower probaility signal PDF tails are less accurately quantised.
One of the possible problem models, where the input signal is first compressed using a so-called non-linear compander characteristic and then uniformly quantised



Stylised non-linear quantiser model using companding



Qualitative effect of companding on PDF shape

If the signal's PDF $p(x)$ is a smooth, known function and sufficiently fine quantisation is used, then the quantisation error variance can be expressed as:

$$\sigma_q^2 \approx \frac{q^2}{12} \int_{-x_{max}}^{x_{max}} \frac{p(x)}{|\dot{c}(x)|^2}\, dx$$

Where $\dot{c}(x) = \frac{dc(x)}{dx}$ represents the slope of the compander's characteristic.
high gradient can de-weights the error contributions due to the highly peaked PDF near the origin

正 匡磊

No.

Date

Minimum quantization error variance is achieved by the compander characteristic

$$C(x) = x_{max} \cdot \frac{\int_0^{x} \sqrt[3]{p(x)}\,dx}{\int_0^{x_{max}} \sqrt[3]{p(x)}\,dx}$$

If the input signal's PDF or variance is time-variant, the compander's performance degrades

§ PDF-independent Quantisation using Logarithmic Compression

The input signal's variance is given in case of an arbitrary PDF $p(x)$ as follows

$$\sigma_x^2 = \int_{-\infty}^{+\infty} x^2 \cdot p(x)\,dx$$

Assume zero saturation distortion, the SNR can be expressed :

$$SNR = \frac{\sigma_x^2}{\sigma_q^2} = \frac{\int_{-x_{max}}^{x_{max}} x^2 p(x)\,dx}{\frac{q^2}{12} \cdot \int_{-x_{max}}^{x_{max}} \left(p(x) / |\dot{C}(x)|^2 \right) dx}$$

In order to maintain an SNR value that is independent of the signal's PDF $p(x)$ the numerator must be a constant times the denominators

$$|\dot{C}(x)|^2 \doteq \left|\frac{K}{x}\right|^2 \quad \text{or} \quad \dot{C}(x) = \frac{K}{x} \leftarrow \text{constant}$$

And hence : $$C(x) = \int_0^x \frac{K}{z}\,dz = K \cdot \ln x + A$$

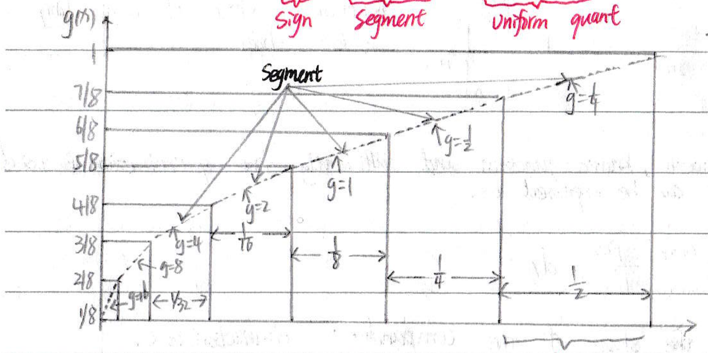
$y = C(x)$, x

This compander ensures a constant SNR across the signal's dynamic range, irrespective of PDF

Recommendation for the transmission of speech sampled at 8 kHz. The logarithmic characteristic is implemented as a 16-segment piece-wise linear approximation

$\underbrace{b_7}_{\text{sign}}$ $\underbrace{b_6 b_5 b_4}_{\text{segment}}$ $\underbrace{b_3 b_2 b_1 b_0}_{\text{uniform quant}}$

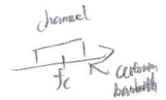
Transmission rate $8 \times 8 = 64$ kbps



divided into 16 uniformly spaced quantisation intervals

No.

Date



No.

Date

No.

Date

2016/10/17

Source Coding Visit

- Transmit at certain rate R_b requires certain amount of resource $\begin{cases} \text{bandwidth} \\ \text{power} \end{cases}$
  - The larger the rate R_b, the larger the required source
- Source coding aims to get R_b as small as possible, ideally close to the source information rate
  - Information rate, a fundamental physical quantity of the source, tells you how many bits/s of information the source really needs to send out.
  - If R_b is close to information rate, source coding is most efficient.
- Memoryless source $\{S(k)\}$, entropy coding on $\{S(k)\}$ is most efficient, as data rate R_b is as to close to source information rate as possible
- For memory source $\{S(k)\}$, information theory also tells us how to code $\{S(k)\}$ most efficiently

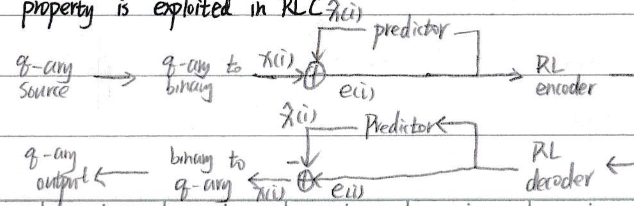
Remove Redundancy

- Key to get close to information rate $H \cdot R_s$ is to remove redundancy.
  - Part of $S(k)$ can be predicted by $\{S(k-i)\}, i \geq 1$ → when coding $S(k)$ this temporal redundancy can be predicted from $\{S(k-i)\}, i \geq 1$
  - By removing the predictable part, resulting residual sequence $\{\varepsilon(k)\}$ is near independent or uncorrelated, can thus be coded by an entropy coding.

- Speech samples are highly correlated (containing lots of redundancy)
  - Predictive speech coding builds a predictive model like $S(k) = \sum_{j=1}^{q} a_j \cdot S(k-j) + \varepsilon(k)$
    $\varepsilon(k)$ contains new information unpredictable from $S(k-j)$, $1 \leq j \leq q$
  - $\{\varepsilon(k)\}$ can then be coded with far smaller rate, and you send $\{\varepsilon(k)\}$ together with the model parameters a_j, $1 \leq j \leq q$, not $\{s(k)\}$

Predictive Scheme with Run-Length Coding (RLC)

- Following predictive scheme exploits partial predictability of memory source:
  - Convert q-ary source to binary sequence $\{x(i)\}$ by BCD
  - Build a predictor $\{\hat{x}(i)\}$ for it
  - If prediction is successful, the "residual" binary sequence $\{e(i)\}$ mostly contains zero, and this property is exploited in RLC



This scheme widely used in video coding, by removing inter-frame and intra-frame correlations as much as possible, resulting binary sequence contains most zeros

"通过前值预测当前值,录像中前后帧变化很小,相减后产生大量的0"

RLC then compresses it into another much shorter binary sequence for transmission

Run Length Coding Table

Code words with fixed length of n bits are formed from a bit stream (encoder input pattern) of upto $L \le N-1 = 2^n - 2$ successive zeros followed by a one or zero.

length of 0-run L	encoder input pattern (length = min $(N, L+1)$)	encoder output codeword (fixed n bits)
0	1	00..000
1	01	00..001
2	001	00…010
3	0001	00…011
⋮	⋮	⋮
$N-2$	0…01	11…101
$N-1$	00…01	11…110
$N=2^n-1$	00…00	11…111

Assumption is input bit stream contains mostly "0"s, i.e. $p = P(\text{"0"})$ is very high

Thus encoder on average reduces the word length

RLC Efficiency

- Code length after run length coding : n bits
- Average code word length d ~~after~~ before coding with $N = 2^n - 1$

$$d = \sum_{L=0}^{N-1} (L+1)\cdot p^L \cdot (1-p) + N \cdot p^N = 1 + p + p^2 + \cdots + p^{N-1} = \frac{1-p^N}{1-p}$$ { P is the probability of a bit is '0' }

- Therefore compression ratio $C = \frac{d}{n}$
- A numerical example : $p = 0.95$, $n = 5$ $(N = 31)$

$$C = \frac{d}{n} = \frac{1-p^N}{n(1-p)} = \frac{\cancel{14.}15.92}{5} \approx 3.18$$

RLC Re-exam Again

Input patterns have variable length, $2^n - 1$ bits to just 1 bit, depending on length of "0" runs before "1" ; while output codewords have fixed length of n bits

[Input pattern]

00…0000	00…0001	00…001	…	01	1
2^n+0 bits	2^n-2+1 bits	2^n-3+1 bits		1+1 bits	0+1 bits

⇓ RLC

[Output pattern]

11…111	11…110	11…101	…	00…001	00…000
n bits	n bits	n bits		n bits	n bits

Shannon-Fano and Huffman : inputs have fixed length while outputs variable lengths

RLC appers very different from Shannon-Fano and Huffman or is it?

All of these three encodings are lossless or entropy encodings

Lossless Encodings Comparison

↳ Input → $\boxed{\text{Encoder}}$ → Output ① Shannon-Fano ② Huffman ③ RLC (for binary data most 0s)

Same principle: rare input pattern/message/symbol coded with large output codeword
large probability coded with small codeword

Shannon-Fano & Huffman : input fixed length → Output variable length

RLC : input variable length → output fixed length

It is the ratio :

$$\text{ratio} = \frac{\text{output length}}{\text{input length}}$$

Small probability → large ratio

large probability → small ratio

Summary

According to information theorey, key to efficient source coding for digital source with memory is first to move the redundancy (then apply entropy coding to "residuals") RESIDAULS

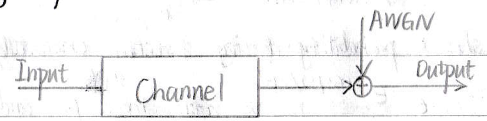
- Efficient coding means coding rate R_b close to information rate $H \cdot R_s$
- Predictable part of current $S(k)$ can be constructed and removed from $S(k)$
- Resulting "residual" sequence $\varepsilon(k) = S(k) - \hat{S}(k)$ is near white, and entropy coding can be applied to it. ~~(then apply~~

Run-length Encoding: a lossless or entropy encoding method.

- For binary digital source, where binary sequence contains most zeros $P(0) \to 1$, $P(1) \to 0$
- RLC compresses such a binary sequence into a much shorter binary sequence
- Compression ration of RLC $c = \frac{d}{n} = \frac{1-p^n}{n(1-p)}$
- Compression with Shannon-Fano and Huffman entropy encoding methods

Information Across Channels

For conveying information across a transmission channel, a suitable model is:



The channel itself introduces amplitude and phase ~~Ha~~ distortion, is potentially time varying, and has a limited bandwidth B.

Error-free reception of symbols is additionally impeded (阻碍) by additive white Gaussian noise (加性高斯白噪声) (AWGN); Its severeness is described by the signal-to-noise ratio (SNR)

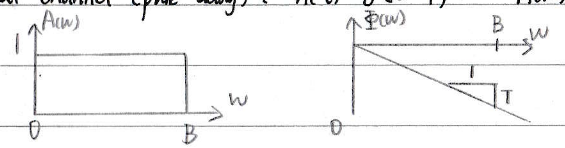
Therefore, dependent on the above parameters, we are interested in determing the maximum possible error-free information transmission (channel capacity C)

We will see that C depends on B and SNR

Characteristics of Channel

The channel can be described by its impulse response $h(t)$ or equivalently its frequency response $H(jw) = A(jw) \cdot e^{j\Phi(w)}$ with Amplitude response $A(jw)$ and phase response $\Phi(w)$; $h(t)$ and $H(jw)$ are Fourier Pair

Ideal channel (pure delay): $h(t) = \delta(t-T) \to A(w) = 1$, $\Phi(w) = -wT$



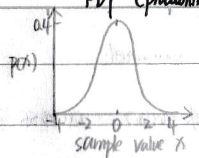
- Flat magnitude and linear phase (= constant group delay $G(w) = -\partial\Phi / \partial(w)$)
- The only impairment (损害) caused by an ideal channel is AWGN

Non-ideal channel: channel is dispersive (分散的), causing intersymbol interfuce

$R(\tau) = \int_{-\infty}^{+\infty} s^*(t)\cdot s(t+\tau)\,dt$ 能量信号自相关函数

当$\tau=0$时，为信号的能量 $R(0) = \int_{-\infty}^{+\infty} s^2(t)\,dt = E$

PDF (probability density function)



Additive White Gaussian Noise

Noise is uncorrelated with the signal

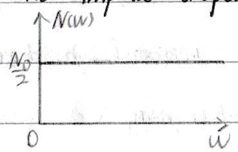
Gaussian noise has a bell shaped probability density function (normally distributed)

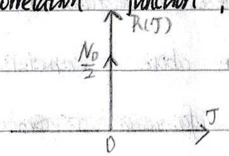
$p(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot e^{-(x-\mu)/2\sigma^2}$ with mean μ (均值) and variance σ^2 (方差)

White noise has zero mean, and channel noise is usually modelled as an AWGN

White Noise

White noise is characterised by a flat power spectral density (功率谱密度函数) function, $N(\omega)$, or equivalently, its impulse-shaped auto-correlation (自相关方程) function, $R(\tau)$





$N(\omega)$ and $R(\tau)$ are a Fourier Pair

$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} N(\omega)\, e^{j\omega\tau} d\omega$ $\quad N(\omega) = \int_{-\infty}^{+\infty} R(\tau)\, e^{-j\omega\tau} d\tau$

Two-side spectrum is usually used for convenience, and No is the noise power

Physic Basis of Channel

Transmitted signal is amplified to required power level and launched from transmit antenna to channel.

- Signal power is attenuated, as it travels in distance - path loss
- Copies of signal arrive at receiver with different attenuation and delays, which may cause dispersive and fading (power level fluctuates (波动) rapidly) effects

Received signal at receiver is very weak, and needs to be amplified to required power level in order to detect ~~signal~~ digital information contained.

- While receiver amplifier amplifies receive signal, it also introduces thermal noise
- AWGN in our channel model in fact models this noise.
- How much noise introduced by amplifier is specified by its noise figure.

Depending on communication carrier frequency, channel bandwidth and actual communication conditions, channel may be modelled as:

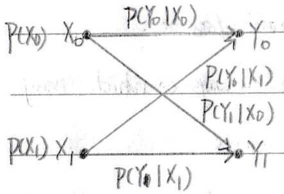
- AWGN channel, i.e. channel is nondispersive or memoryless
- Or dispersive channel, i.e. channel has memory

$H(X) = 0.8812$

0.891

Binary Symmetric Channel (BSC)

BSC is the simplest model for information transmission via a discrete channel (channel is ideal, no amplitude and phase distortion, only distortion is due to AWGN):



$P(X_i)$: probability of occurance of X_i at source output, a prior probability

$P(Y_j|X_i)$: conditional probability of R_x Y_j given T_x X_i

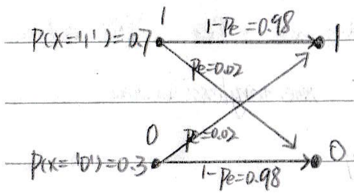
"Symmetric" refers to $P(Y_1|X_0) = P(Y_0|X_1) = P_e$ ← Channel Error Probability

The joint probability $P(Y_j, X_i)$ (T_x X_i and R_x Y_j) is linked with the conditional probability $P(Y_j|X_i)$ by Bayes' rule:

$$\boxed{P(Y_j, X_i) = P(X_i)\cdot P(Y_j|X_i) = P(Y_j)\cdot P(X_i|Y_j) = P(X_i, Y_j)}$$

Binary Symmetric Channel - Example

Consider a BSC:



- This is a non-equiprobable source with $P(X='1')=0.7$ and $P(X='0') = 0.3$

on average, 70% of transmitted bits are "1" and 30% are "0"

- Channel's error probability $P_e = 0.02$: on average, bit error rate 2%

{ X|Y → Y是X发生的前提条件
{ X, Y → X与Y同时发生

Probability of correct reception: $P_{correct} = P(Y='1', X='1') + P(Y='0', X='0') = 0.98$

$$P(Y='1', X='1') = P(X='1')\cdot P(Y='1'|X='1') = 0.7\cdot 0.98 = 0.686$$

$$P(Y='0', X='0') = P(X='0')\cdot P(Y='0'|X='0') = 0.3\cdot 0.98 = 0.294$$

Probability of erroneous reception: $P_{error} = P(Y='1', X='0') + P(Y='0', X='1') = 0.02$

$$P(Y='1', X='0') = P(X='0')\cdot P(Y='1'|X='0') = 0.3\cdot 0.02 = 0.006$$

$$P(Y='0', X='1') = P(X='1')\cdot P(Y='0'|X='1') = 0.7\cdot 0.02 = 0.014$$

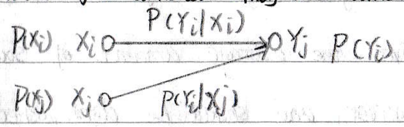
Total probability of receiving a '1' or a '0'

$$P(Y='1') = P(X='1')\cdot P(Y='1'|X='1') + P(X='0')\cdot P(Y='1'|X='0') = 0.7\cdot 0.98 + 0.3\cdot 0.02 = 0.692$$

$$P(Y='0') = P(X='0')\cdot P(Y='0'|X='0') + P(X='1')\cdot P(Y='0'|X='1') = 0.3\cdot 0.98 + 0.7\cdot 0.02 = 0.308$$

A Close Look at Channel

Y_i and X_j are connected → They have something in common or "mutual"



Mutual Information

Definition of mutual information of X_i and Y_j: $\boxed{I(X_i, Y_j) = \log_2 \frac{P(X_i|Y_j)}{P(X_i)}}$?

Perfect, noiseless channel: $Y_i = X_i$, i.e. $P(X_i|Y_i) = 1$ and $I(X_i, Y_i) = \log_2 \frac{1}{P(X_i)}$

- This is the information of X_i, hence no information is lost in the channel

Extremely noisy channel with error probability 0.5 → Y_i is independent of X_i

$$P(X_i|Y_i) = \frac{P(X_i, Y_i)}{P(Y_i)} = \frac{P(X_i)\cdot P(Y_i)}{P(Y_i)} \qquad \frac{P(X_i|Y_j)}{P(X_i)} = 1 \Rightarrow \log_2 1 = 0$$

- Therefore $I(X_i, Y_i) = \log_2 1 = 0$, meaning all information all information is lost in the channel

In general, $I(X_i) > I(X_i, Y_i)$, some information is lost in the channel

Mutual Information - Example

$P(X_1|Y_1) = 0.9913 = P(Y_1, X_1)/P(Y_1) = 0.686/0.692 = 0.9913$

$P(X_0|Y_1) = 0.0087 = P(Y_1, X_0)/P(Y_1) = 0.006/0.692 = 0.0087$

$P(X_1|Y_0) = 0.0455 = P(Y_0, X_1)/P(Y_0) = 0.014/0.308 = 0.0455$

$P(X_0|Y_0) = 0.9545 = P(Y_0, X_0)/P(Y_0) = 0.294/0.308 = 0.9545$

$I(X_1, Y_1) = \log_2(P(X_1|Y_1)/P(X_1)) = 0.502$ bits $\qquad I(X_0, Y_0) = 1.670$ bits

(Source Info: $I(X_1) = \log_2(1/P(X_1)) = 0.515$ bits $\qquad I(X_0) = 1.737$ bits $= \log_2(1/P(X_0))$

(Destin Info: $I(Y_1) = \log_2(1/P(Y_1)) = 0.531$ bits $\qquad I(Y_0) = 1.699$ bits $= \log_2(1/P(Y_0))$

$I(X_0, Y_1) = \log_2(P(X_0|Y_1)/P(X_0)) = \log_2 \frac{0.0087}{0.3} = -5.113$ bits $\qquad I(X_1, Y_0) = -3.945$ bits

- Destination info contents are more balanced ~~when~~ then source info contents
- The negative quantities represent "mis-information"

No.

Date

Average Mutual Information

Based on received symbols Y_j given transmitted symbols X_i through a BSC, average mutual information is defined as:

$$I(X,Y) = \sum_i \sum_j P(X_i, Y_j) \cdot I(X_i, Y_j) = \sum_i \sum_j P(X_i, Y_j) \cdot \log_2 \frac{P(X_i|Y_j)}{P(X_i)}$$

This gives the average amount of source information acquired per received symbol by the receiver, and should be distinguished from the average source information (entropy $H(X)$)

Note that due to Bayes: $\frac{P(X_i|Y_j)}{P(X_i)} = \frac{P(X_i, Y_j)}{P(X_i) \cdot P(Y_j)} = \frac{P(Y_j|X_i)}{P(Y_j)}$

Imperfect Channel : Information Loss

Consider re-arranging the mutual information between transmitted symbol X_i and received symbol Y_j

$$I(X_i, Y_j) = \log_2 \frac{P(X_i|Y_j)}{P(X_i)} = \log_2 \frac{1}{P(X_i)} - \log_2 \frac{1}{P(X_i|Y_j)} = I(X_i) - I(X_i|Y_j)$$

$I(X_i, Y_j)$ is the amount of information conveyed to receiver when transmitting X_i and receiving Y_j, $I(X_i)$ is the source information of X_i, and $I(X_i|Y_j)$ can be regarded as the information loss due to the channel.

Therefore,

$$\underbrace{I(X_i)}_{\text{Source Inf}} - \underbrace{I(X_i, Y_j)}_{\text{Inf conveyed to rec.}} = \underbrace{I(X_i|Y_j)}_{\text{Inf. loss}}$$

$\log_2 \frac{1}{P(X_i)} - I(X_i;Y_j) = \log_2 \frac{1}{P(X_i|Y_j)}$

$I(X_i;Y_j) = \log_2 \frac{P(X_i|Y_j)}{P(X_i)}$

$0 \le I(X_i|Y_j) \le I(X_i)$

No.

Date

Imperfect Channel : Average Mutual Information

$$I(X,Y) = \sum_i \sum_j P(X_i, Y_j) \cdot \log_2 \frac{P(X_i|Y_j)}{P(X_i)} \quad \text{[bits/symbol]}$$

Average Mutual Information

- But this average conveyed information

$$I(X,Y) = \sum_i \sum_j P(X_i, Y_j) \cdot \log_2 \frac{1}{P(X_i)} - \sum_i \sum_j P(X_i, Y_j) \cdot \log_2 \frac{1}{P(X_i|Y_j)}$$

$P(X_i, Y_j) = P(Y_j) \cdot P(X_i|Y_j)$

$$= \sum_i \left(\sum_j P(X_i, Y_j) \right) \cdot \log_2 \frac{1}{P(X_i)} - \sum_j P(Y_j) \cdot \left(\sum_i P(X_i|Y_j) \cdot \log_2 \frac{1}{P(X_i|Y_j)} \right)$$

$$= \sum_i P(X_i) \cdot \log_2 \frac{1}{P(X_i)} - \sum_j P(Y_j) \cdot I(X|Y_j)$$

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

$$\underbrace{I(X,Y)}_{\text{average conveyed information}} = \underbrace{H(X)}_{\text{average source information}} - \underbrace{H(X|Y)}_{\text{average information lost}}$$

Source entropy

A similar re-arrangement leads to: $\underbrace{I(X,Y)}_{\text{average conveyed information}} = \underbrace{H(Y)}_{\text{~~average source information~~ destination}} - \underbrace{H(Y|X)}_{\text{error entropy}}$

error

$$I(X,Y) = \sum_i \sum_j P(Y_j, X_i) \cdot \log_2 \frac{P(Y_j|X_i)}{P(Y_j)}$$

$P(Y_j, X_i) = P(X_i) \cdot P(Y_j|X_i)$

$$= \sum_i \sum_j P(Y_j, X_i) \cdot \log_2 \frac{1}{P(Y_j)} - \sum_i \sum_j P(Y_j, X_i) \log_2 \frac{1}{P(Y_j|X_i)}$$

$$= \sum_i \left(\sum_j P(Y_j, X_i) \log_2 \frac{1}{P(Y_j)} \right) - \sum_i P(X_i) \cdot \sum_j P(Y_j|X_i) \cdot \log_2 \frac{1}{P(Y_j|X_i)}$$

$$= \sum_i P(Y) \cdot \log_2 \frac{1}{P(Y_j)} - \sum_i P(X_i) \cdot I(Y|X_i)$$

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

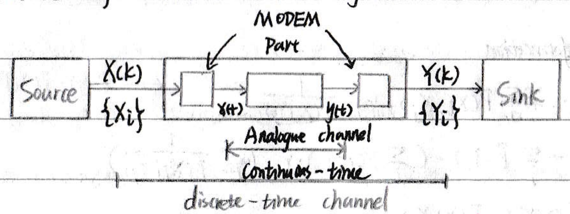
No.

Date

InfThe - L6

A Closer Look at Channel

Schematic of communication system



- Depending which of ~~system~~ part of system : discrete-time & continuous-time channels

Source has information rate, we would like to know "capacity" of channel

- According to Shannon : information rate ≤ capacity → error-free trasfer

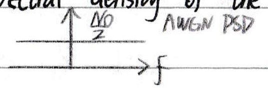
Review of Channel Assumption

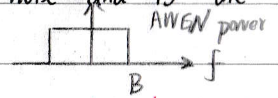
No amplitude or phase distortion by the channel, and the only distortion is due to additive white Gaussian noise (AWGN), i.e. ideal channel.

- In the simplest case, this can be modelled by a binary symmetric channel (BSC)

The channel error probability P_e of the BSC depends on the noise power N_p relative to the signal power S_p, i.e. $SNR = \frac{S_p}{N_p}$

- Hence P_e could be made arbitrarily small by increasing the signal power
- The channel noise power can be shown to be $N_p = N_0 \cdot B$, where $\frac{N_0}{2}$ is the power spectral density of the noise and B the channel Bandwidth





Our aim is to determine the channel capacity C, the maximum possible error-free information transmission rate across the channel

No.

Date

Channel Capacity for Discrete Channels

Shannon's channel capacity C is based on the average mutual information (average conveyed information across the channel), and one possible definition is:

$$C = \max\{I(X,Y)\} = \max\{H(Y) - H(Y|X)\} \quad [\text{bits/symbol}]$$

where H(Y) is the average information per symbol at channel ~~ott~~ output or destination entropy, and H(Y|X) error entropy.

Let t_i be the symbol duration for X_i and t_{av} be the average time for transmission of a symbol, the channel capacity can also be defined as

$$C = \max\{I(X,Y)/t_{av}\} \quad [\text{bits/symbol}]$$

C becomes maximum if $H(Y|X) = 0$ (no errors) and the symbols are equiprobable (assuming constant symbol durations t_i)

Channel capacity can be expressed in either (bits/symbol) or (bits/second)

Channel Capacity : Noise-Free Case

In noise free case, error entropy $H(Y|X) = 0$ and $I(X,Y) = H(Y) = H(X)$

- But the entropy of the source is given by:

$$H(X) = -\sum_{i=1}^{q} P(X_i) \cdot \log_2 P(X_i) \quad [\text{bits/symbol}]$$

- Let t_i be symbol duration for X_i; average time for transmission of a symbol is

$$t_{av} = \sum_{i=1}^{q} P(X_i) \cdot t_i \quad [\text{seconds/symbol}]$$

- By definition, the channel capacity is $C = \max\{H(X)/t_{av}\}$ [bits/second]

No.

Date

Assuming constant symbol duration $t_i = T_s$, the maximum or the capacity is obtained for memoryless q-ary with equiprobable symbols.

$$C = \log_2 q / T_s$$

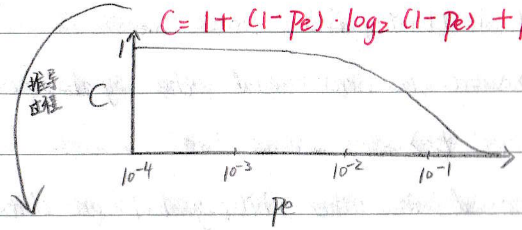
- This is the maximum achievable information transmission rate

Channel Capacity for BSC

BSC with equiprobable source symbols $P(X_0) = P(X_1) = 0.5$ and variable channel error probability P_e (due to symmetry of BSC, $P(Y_0) = P(Y_1) = 0.5$)

The channel capacity C (in bits/symbol) is given as

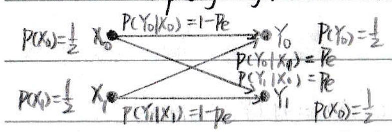
$$C = 1 + (1-P_e)\cdot\log_2(1-P_e) + P_e\cdot\log_2 P_e$$



$P_e = 0.5$ (worst case) $\Rightarrow C = 0$

$P_e = 0$ (best case) $\Rightarrow C = 1$

Channel Capacity for BSC (Derivation)



$$P(X_0,Y_0) = P(X_0)\cdot P(Y_0|X_0) = (1-P_e)/2$$
$$P(X_0,Y_1) = P(X_0)\cdot P(Y_1|X_0) = P_e/2$$
$$P(X_1,Y_1) = (1-P_e)/2$$
$$P(X_1,Y_0) = P_e/2$$

$$I(X,Y) = \sum_i\sum_j P(X_i,Y_j)\cdot\log_2\frac{P(X_i|Y_j)}{P(X_i)} = P(X_0,Y_0)\cdot\log_2\frac{P(X_0|Y_0)}{P(X_0)} + P(X_0,Y_1)\cdot\log_2\frac{P(X_0|Y_1)}{P(X_0)}$$

$P(X_0)\cdot P(Y_0|X_0) = P(Y_0)\cdot P(X_0|Y_0)$

$$+ P(X_1,Y_0)\cdot\log_2\frac{P(X_1|Y_0)}{P(X_1)} + P(X_1,Y_1)\cdot\log_2\frac{P(X_1|Y_1)}{P(X_1)}$$

$$= \frac{1-P_e}{2}\log_2 2(1-P_e) + \frac{P_e}{2}\cdot\log_2 2P_e + \frac{P_e}{2}\log_2 2P_e + \frac{1-P_e}{2}\log_2 2(1-P_e)$$

$$= P_e\log_2 2P_e + (1-P_e)\cdot\log_2 2(1-P_e)$$

$$= P_e\log_2 2 + P_e\log_2 P_e + (1-P_e)\cdot\log_2 2 + (1-P_e)\cdot\log_2(1-P_e)$$
$$= P_e + P_e\log_2 P_e + 1 - P_e + (1-P_e)\cdot\log_2(1-P_e)$$

$$= 1 + (1-P_e)\cdot\log_2(1-P_e) + P_e\cdot\log_2 P_e$$

No.

Date

Channel Capacity and Channel Coding

Shannon's theorem: If information rate $R \le C$, there exists a coding technique such that information can be transmitted over the channel with arbitrarily small error probability; if $R > C$, error-free transmission is impossible.

- C is the maximum possible error-free information transmission rate
- Even in noisy channel, there is no obstruction (阻碍) of reliable transmission, but only a limitation of the rate at which transmission can take place.

Most practical channel coding schemes are far from optimal, but capacity-approaching codes exist. e.g. $\boxed{\text{turbo codes}}$ and $\boxed{\text{low-density parity check codes}}$

Practical communication systems are far from near capacity, but recently near-capacity techniques have been developed - iterative turbo detection - decoding

Entropy of Analogue Source

Entropy of a continuous-valued (analogue) source, where the source output x is described by the PDF $p(x)$, is defined by

$$H(x) = -\int_{-\infty}^{+\infty} p(x)\cdot\log_2 p(x)\,dx$$

According to Shannon, this entropy attends the maximum for Gaussian PDFs $p(x)$ (equivalent to equiprobable symbols in the discrete case)

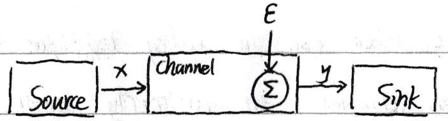
Gaussian PDF with zero mean and variance σ_x^2:

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-(x^2/2\sigma_x^2)}$$

The maximum entropy can be shown to be

$$H_{max}(x) = \log_2\sqrt{2\pi e}\,\sigma_x = \frac{1}{2}\log_2 2\pi e\sigma_x^2$$

Gaussian Channel



Signal with Gaussian PDF attains maximum entropy, thus we consider Gaussian channel

Channel output y is linked to channel input x by

$$y = x + \varepsilon$$

- Channel AWGN ε is independent of channel input x, having noise power $N_p = \sigma_\varepsilon^2$
- Assume Gaussian channel, i.e. channel input x has a Gaussian PDF, having signal power $S_p = \sigma_x^2$
- Basic linear system theory: Gaussian signal operated by linear operator remains Gaussian

Channel output signal y is also Gaussian, i.e. having a Gaussian PDF with power σ_y^2

$$\sigma_y^2 = S_p + N_p = \sigma_x^2 + \sigma_\varepsilon^2$$

Thus channel output y has a Gaussian PDF

$$p(y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-(y^2/2\sigma_y^2)}$$

with power $\sigma_y^2 = S_p + N_p$, and entropy of y attains maximum value

$$H_{max}(y) = \frac{1}{2}\log_2 2\pi e\,(S_p + N_p)$$

Since $I(x,y) = H(y) - H(y|x)$, and $H(y|x) = H(\varepsilon)$ with ε being AWGN

$$H(y|x) = \frac{1}{2}\log_2 2\pi e N_p$$

Therefore, the average mutual information

$$I(x,y) = \frac{1}{2}\log_2\left(1 + \frac{S_p}{N_p}\right) \text{ [bits/symbol]}$$

Shannon - Hartley Law

With a sampling rate of $f_s = 2\cdot B$, the Gaussian channel capacity is given by

$$C = f_s \cdot I(x,y) = B\cdot \log_2\left(1 + \frac{S_p}{N_p}\right) \text{ [bits/second]}$$

where B is the signal bandwidth

- For digital communications, signal bandwidth B (Hz) is channel bandwidth
- Sampling rate f_s is the symbol rate (symbol/second)
- Channel noise power is $N_p = N_0 \cdot B$, where N_0 is the power spectral density of the channel AWGN

Two basic resources of communication: bandwidth and signal power

- Increasing the SNR $\frac{S_p}{N_p}$ increases the channel capacity
- Increasing the channel bandwidth B increases the channel capacity

Gaussian channel capacity is often a good approximation for practical digital communication channels

Bandwidth and SNR Trade-off

From the defination of channel capacity, we can trade the channel bandwidth B for the SNR or signal power S_p, and vice versa.

- Depending on whether B or S_p is more precious, we can increase one and reduce the other, and yet maintain the same channel capacity.
- A noiseless analogue channel ($\frac{S_p}{N_p} = \infty$) has an infinite capacity

C increases as B increases, but it does not go to infinity as $B \to \infty$, rather C approaches an upper limit

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

$$\lim_{x\to 0}(1+x)^{\frac{1}{x}} = e \Rightarrow C_\infty = \lim_{B\to\infty} C = \frac{S_p}{N_0}\cdot\log_2 e = 1.44\frac{S_p}{N_0}$$

Bandwidth and SNR Trade off - Example

Q: A channel has an SNR of 15. If the channel bandwidth is reduced by half, determine the increase in the signal power required to maintain the same channel capacity.

$SNR = \frac{S_p}{N_p} = \frac{S_p}{N_0 B} = 15$

A: $B \cdot \log_2(1 + \frac{S_p}{N_0 B}) = B' \cdot \log_2(1 + \frac{S_p'}{N_0 B'})$

$B \cdot \log_2 16$

$4 \cdot B = \frac{B}{2} \cdot \log_2(1 + \frac{(S_p'/S_p) \cdot S_p}{N_0 B/2})$

$8 = \log_2(1 + 30\frac{S_p'}{S_p})$

2^8

$256 = 1 + 30\frac{S_p'}{S_p} \Rightarrow S_p' = 8.5 S_p$

Summary

Channel capacity, a fundamental physical quantity, defines maximum rate that information can be transfered across channel error-free

- It is based on concept of maximum achievable mutual information between channel input and output, either defined as [bits/symbol] or [bits/s]
- Channel capacity for discrete channels, eg channel capacity for BSC
- Shannon theorem

Channel capacity for continuous channels

- Continuous-valued signal attains maximum entropy, if its PDF is Gaussian
- Gaussian channel capacity: Shanon-Hartley Law

$C = B \cdot \log_2(1 + \frac{S_p}{S_N})$ [bits/second]

Bandwidth SNR

Information Theory Revision (Source)

digital source $\xrightarrow{\{S(k)\}}$ symbols/s source coding $\xrightarrow{\{b_i\}}$ bits/s

- Digital source is defined by
  1. Symbol set: $S = \{m_i, 1 \le i \le 8\}$
  2. Probability of occuring of m_i: P_i, $1 \le i \le 8$
  3. Symbol rate: R_s [symbols/s]
  4. ~~Indepen~~ Interdependency of $\{S(k)\}$
- Information Content of alphabet m_i: $I(m_i) = -\log_2(P_i)$ [bits]
- Entropy quantifies average information conveyed per symbol
  - Memoryless sources: $H = -\sum_{i=1}^{8} P_i \cdot \log_2(P_i)$ [bits/symbol]
  - 1st-order memory (1st-order Markov) sources with transition probabilities P_{ij}:

$$H = \sum_{i=1}^{8} P_i H_i = -\sum_{i=1}^{8} P_i \sum_{j=1}^{8} P_{ij} \cdot \log_2(P_{ij}) \text{ [bits/symbol]}$$

- Information rate: tells you how many bits/s information the source really needs to send out
  - Information rate: $R = R_s \cdot H$ [bits/s]
- Efficient source coding: get rate R_b as close to information rate R as possible
  - Memoryless source: apply entropy coding, such as Shannon-Fano and Huffman, and RLC if source is binary with most zeros.
  - Generic sources with memory: remove redundancy first, then apply entropy coding to "~~resd~~ residuals"

Practical Source Coding

Practical source coding is guided by information theory, with practical constraints, such as performance and processing complexity / delay trade off.

As we will learn, data rate is directly linked to required bandwidth, source coding is to encode source with a data rate as small as possible $R_b \to R$

For correlated source or signal with memory, to achieve this

1. Critical to remove redundancy or correlation of source signal first.
2. After removing predictable part, resulting residual signal is nearly white, can then be coded with entropy coding, such as Shannon-Fano or Huffman, or RLC

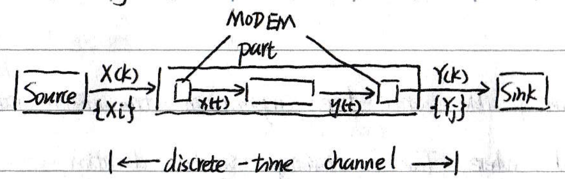
Speech signal has significant temporal correlation

- When you come to speech codecs, you can always identify 1. and 2., and see how temporal redundancy is dealt with in different practical ways.

Video signal has even more correlations: intra-frame (spatial) correlation and inter-frame (temporal) correlation.

- When you come to video codecs, you can always identify 1. and 2., and see how both spatial and temporal redundancies are dealt with in different practical ways.
- e.g. two consecutive video frames differ very little (temporal correlation): send a whole frames as reference, then only send difference of two consecutive frames which may be coded by RLC.

Information Theory Revision (Channel)



Information theory studies what happens to information transferring across channel, and provides guiding principles for designing of practical communication systems.

Average mutual information $I(X,Y)$ between channel input $X(k) \in \{X_i\}$ and channel output $Y(k) \in \{Y_j\}$ charaterises how information is transferring across channel.

$$I(X,Y) = \sum_i \sum_j P(X_i, Y_j) \cdot \log_2 \frac{P(X_i|Y_j)}{P(X_i)} \quad [\text{bits/symbol}]$$

- What happened to information transferred across channel:

$$\underbrace{I(X,Y)}_{\text{av. conveyed inf}} = \underbrace{H(X)}_{\text{source entropy}} - \underbrace{H(X|Y)}_{\text{av. information lost}}$$

$$\underbrace{I(X,Y)}_{\text{av. conveyed inf}} = \underbrace{H(Y)}_{\text{destination entropy}} - \underbrace{H(Y|X)}_{\text{error entropy}}$$

Channel Capacity

- Channel capacity C is maximum possible error-free information transmission rate across channel.
- Channel capacity of discrete channel, where T_{av} is average symbol duration.

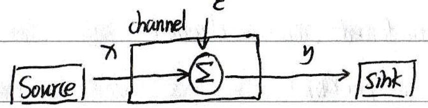
$$C = \max I(X,Y) \text{ [bits/symbol] or } C = \max I(X,Y)/T_{av} \text{ [bits/s]}$$

Channel capacity of BSC with error probability P_e

$$C = 1 + (1-P_e)\log_2(1-P_e) + P_e \log_2 P_e \text{ [bits/symbol]}$$

Best case: $P_e = 0$ and $C = 1$ [bits/symbol] ; Worst case: $P_e = 0.5$ and $C = 0$ [bits/symbol]

- Gaussian channel capacity and Shannon-Hartley expression



$$C = B \cdot \log_2\left(1 + \frac{S_p}{N_p}\right) \text{ [bits/s]}$$

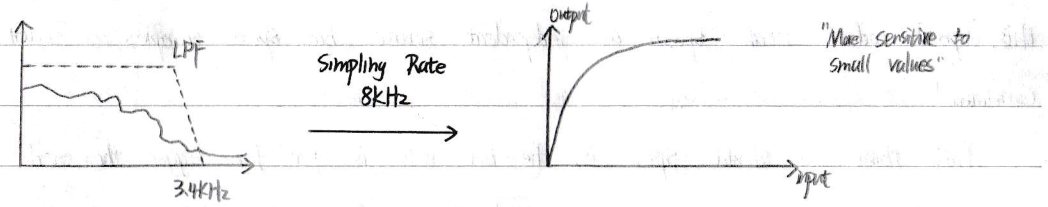
- Two basic resources are channel bandwidth B [Hz] and signal power S_p
- For AWGN ε with two-sider power spectral density $\frac{N_0}{2}$, noise power $N_p = B \cdot N_0$
- Trade of between B and S_p

60's 64kbps "Speech Codec"

When analogue telephone network was digitised, the following 64kpbs "speech codec" was defined

- Analogue speech signal is passed through a low pass filter with cut-off frequency 3.4KHz, and filtered signal is sampled at sampling rate at 8k symbols/s
- As speech signal samples are most have small ~~mang~~ magnitudes and large-magnitudes are rare, samples are passed through a μ-law compressor

Speech Spectrum



- Compressed samples are quantised by a 8-bit quantiser, i.e. quantised to 256 symbols (levels)

Source: alphabet of 8 256 symbols → symbol rate $R_s = 8k$ symbols/s, $\log_2 8 = 8$ bits; actual bit rate: $R_b = 64$ kpbs

Critics

Most of students simply work out this has a bit rate of 64kbps

- Most of you know low pass filtering is to limit speech signal bandwidth, and as most of speech energy is in low frequency band, cut off frequency of 3.4kHz is appropriate.
- As filtered signal is band limited to 3.4KHz, you know you must samples at least twice of this, so sampling rate of 8kHz is appropriate.

Clever ones then "spot" - speech samples (magnitudes) are not equiprobable, and BCD of 8 bits is not right, an entropy encoding should be used - this is actually not correct!

- μ-law compressor compresses large-magnitude samples but is also expands small-magnitude samples, so it should be called compressor-expandor
- Compressed and expanded samples are more near equiprobale - you can see the μ-law compressor is some sort of practical implementation of entropy encoding

So what is wrong ? If speech source is memoryless , then all sounds great , or this "speech codec" treats speech as independent source , but speech samples are HIGHLY Correlated !

- First thing you should spot is the bit rate is far far bigger than ~~actl sounds great~~ really necessary , i.e. much much larger than information rate.
- Remove redundancy , e.g. build a predictive model , after remove predictable part, resulting residual sequence will be near white and can be coded with far far smaller bit rate.
- In practice , you of course need to send your predictor parameters , look at your mobile phone, its speech codec probably only has a few kbps bit rate.

[EXAMP 1]

X_i	$P(X_i)$	BCD word
A	0.30	000
B	0.10	001
C	0.02	010
D	0.15	011
E	0.40	100
F	0.03	101

[1] A source emits symbols X_i , $1 \le i \le 6$, in the BCD format with probabilities $P(X_i)$ as given in Table 1, at a rate $R_s = 9.6$ kbaud [Symbol/second]

(i) State the information rate

$$H = \sum_i^6 P(X_i) \cdot \log_2 \frac{1}{P(X_i)} = 2.05724 \text{[bits/symbol]}$$

$$R_b = H \cdot R_s = 9600 \times 2.05724 = 19750 \text{[bits/s]}$$

(ii) The data rate of the source

$$\text{Data Rate} = 9600 \cdot 3 = 28800 \text{ [bits/s]}$$

[2] Apply Shannon-Fano Coding to the source characterised in Table . Are there any disadvantages in the resulting code words

E	0.4	0				
A	0.3	1	0			
D	0.15	1	1	0		
B	0.1	1	1	1	0	
F	0.03	1	1	1	1	0
C	0.02	1	1	1	1	1

Disadvantage : the rare code words have maximum possible length of $q-1=5$, and a buffer of 5 bits is required, compared with 3 bits for BCD

[3] What is the original symbol sequence of the Shannon-Fano coded signal

D E F E E E D A D E

110|0|11110|0|0|0|110|10|110|0|

[4] What is the data rate of the signal after Shannon-Fano coding ?

What compression factor has been achieved ?

$$I_{av} = 0.4 \times 1 + 0.3 \times 2 + 0.15 \times 3 + 0.1 \times 4 + 0.03 \times 5 + 0.02 \times 5 = 2.1 \text{ [bits/symbol]}$$

$$R_b = I_{av} \cdot R_s = 2.1 \cdot 9600 = 20160 \text{ [bits/s]}$$

$$\text{Compress Ratio} = \frac{3}{2.1} = 1.4286$$

[5] Derive the coding efficiency of both uncoded BCD signal as well as Shannon-Fano coded Signal

$$CE_{BCD} = \frac{R}{R_s} = \frac{19750}{28800} = 68.58\%$$

$$CE_{Shannon} = \frac{R}{R_s} = \frac{19750}{20160} = 97.97\%$$

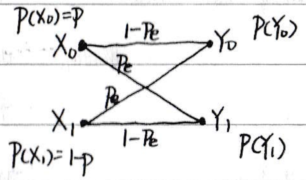
[6] Repeat part 2 to 5 but this time with Huffman coding

EE A EEEE A A EE D EE A
1|1|00|1|1|1|1|00|00|1|1|010|1|1|00|

X_i	$P(X_i)$	Steps	Code
E	0.4	1	1
A	0.3	0 0	00
D	0.15	0 1 0	010
B	0.1	0 1 1 0	0110
F	0.03	0 1 1 1 0	01110
C	0.02	1 1 1 1 0	01111

[EXAMPLE 2]

[i] Consider a binary symmetric channel (BSC)



Mutual information

$$I(X,Y) = \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{P(X_i|Y_j)}{P(X_i)} \quad [\text{bits/symbol}]$$

$H(X) = \sum_j P(Y_j)\cdot P(X_i)\cdot \log_2 \frac{1}{P}$

(i) a formula relating $I(X,Y)$, the source entropy $H(X)$, and the average information lost per symbol $H(X|Y)$

$$I(X,Y) = \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{P(X_i|Y_j)}{P(X_i)} = \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{1}{P(X_i)} - \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{1}{P(X_i|Y_j)}$$

$$= \sum_i P(X_i)\cdot \log_2 \frac{1}{P(X_i)} - \sum_j P(Y_j) \sum_i P(X_i|Y_j)\cdot \log_2 \frac{1}{P(X_i|Y_j)}$$

$$= H(X) - \sum_j P(Y_j)\cdot I(X|Y_j) = H(X) - H(X|Y)$$

(ii) a formula relating $I(X,Y)$, the destination entropy $H(Y)$, and the error entropy $H(Y|X)$.

$$I(X,Y) = \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{P(X_i|Y_j)}{P(X_i)} = \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{P(Y_j|X_i)}{P(Y_j)} = \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{1}{P(Y_j)} - \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{1}{P(Y_j|X_i)}$$

$$= \sum_j \left(\sum_i P(X_i,Y_j)\cdot \log_2 \frac{1}{P(Y_j)}\right) - \sum_i P(X_i) \sum_j P(Y_j|X_i)\cdot \log_2 \frac{1}{P(Y_j|X_i)} = \sum_j P(Y_j)\cdot \log_2 \frac{1}{P(Y_j)} - \sum_i P(X_i)\, I(Y|X_i)$$

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

2. State and justify the relation ($>, <, =, \leq$, or $\geq$) between $H(X|Y)$ and $H(Y|X)$

Unless $P_e = 0.5$ or for equiprobable source symbols X, the symbols Y at the destination are more balance, hence $H(Y) \geq H(X)$. Therefore, $H(Y|X) \geq H(X|Y)$

3. Considering the BSC, we now have $p = \frac{1}{4}$ and a channel error probability $P_e = \frac{1}{10}$

Calculate all probabilities $P(X_i, Y_j)$ and $P(X_i|Y_j)$, and derive the numerical value for the mutual information $I(X,Y)$.

$P(X_0) = \frac{1}{4}$ $\quad P(Y_0|X_0) = \frac{9}{10}$ $\quad P(Y_0) = P(X_0)\cdot P(Y_0|X_0) + P(X_1)\cdot P(Y_0|X_1) = \frac{1}{4}\cdot\frac{9}{10} + \frac{3}{4}\cdot\frac{1}{10} = 0.3$

$P(X_1) = \frac{3}{4}$ $\quad P(Y_1|X_0) = \frac{1}{10}$ $\quad P(Y_1) = P(X_0)\cdot P(Y_1|X_0) + P(X_1)\cdot P(Y_1|X_1) = \frac{1}{4}\cdot\frac{1}{10} + \frac{3}{4}\cdot\frac{9}{10} = 0.7$

$P(Y_0|X_1) = \frac{1}{10}$

$P(Y_1|X_1) = \frac{9}{10}$

$P(X_0,Y_0) = P(X_0)\cdot P(Y_0|X_0) = \frac{1}{4}\cdot\frac{9}{10} = 0.225$ $\quad P(X_0|Y_0) = \frac{P(X_0,Y_0)}{P(Y_0)} = 0.75$

$P(X_0,Y_1) = P(X_0)\cdot P(Y_1|X_0) = \frac{1}{4}\cdot\frac{1}{10} = 0.025$ $\quad P(X_0|Y_1) = \frac{P(X_0,Y_1)}{P(Y_1)} = 0.0357$

$P(X_1,Y_0) = P(X_1)\cdot P(Y_0|X_1) = \frac{3}{4}\cdot\frac{1}{10} = 0.075$ $\quad P(X_1|Y_0) = \frac{P(X_1,Y_0)}{P(Y_0)} = 0.25$

$P(X_1,Y_1) = P(X_1)\cdot P(Y_1|X_1) = \frac{3}{4}\cdot\frac{9}{10} = 0.675$ $\quad P(X_1|Y_1) = \frac{P(X_1,Y_1)}{P(Y_1)} = 0.964$

$$I(X,Y) = \sum_i \sum_j P(X_i,Y_j)\cdot \log_2 \frac{P(X_i|Y_j)}{P(X_i)} = P(X_0,Y_0)\cdot \log_2 \frac{P(X_0|Y_0)}{P(X_0)} + P(X_0,Y_1)\cdot \log_2 \frac{P(X_0|Y_1)}{P(X_0)} + P(X_1,Y_0)\cdot \log_2 \frac{P(X_1|Y_0)}{P(X_1)} + P(X_1,Y_1)\cdot \log_2 \frac{P(X_1|Y_1)}{P(X_1)}$$

$$= 0.225\cdot \log_2 \frac{0.75}{0.25} + 0.025\cdot \log_2 \frac{0.0357}{0.25} + 0.075\cdot \log_2 \frac{0.25}{0.75} + 0.675\cdot \log_2 \frac{0.964}{0.75}$$

$$= 0.3566 - 0.070198 - 0.1189 + 0.2444$$

$$= 0.411902 \quad [\text{bits/Symbol}]$$

[EXAMPLE 3]

A digital communication system uses a 4-ary signalling scheme with symbol set $\{x_1=-3, x_2=-1, x_3=1, x_4=3\}$. The channel is an ideal additive white Guassian noise (AWGN) channel, the transmission rate is 100M Baud (10^8 symbols/s) and the channel signal to noise ratio is know to be 63. The probabilities of occurance for the four symbols at transmitter are respectively.

$$P(x_1)=0.2, \quad P(x_2)=0.3, \quad P(x_3)=0.2, \quad P(x_4)=0.3$$

It is known that this 4-ary symbol source is a first order Markov process with the known transition probability matrix

$$T = \cancel{P[X_i]}\ P[p_{ij}] = \begin{bmatrix} 0.6 & 0.1 & 0.1 & 0.2 \\ 0.1 & 0.6 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.8 & 0.0 \\ 0.0 & 0.1 & 0.1 & 0.8 \end{bmatrix} \quad \text{where } p_{ij}=P(x_j|x_i),\ 1\le i, j\le 4$$

[1] Determine the information rate of this 4-ary source.

$$H_1=\sum_{j=1}^{4} P_{1j}\cdot\log_2\frac{1}{P_{1j}} = H_2=\sum_{j=1}^{4} P_{2j}\cdot\log_2\frac{1}{P_{2j}} = 1.571\ [\text{bits/symbol}]$$

$$H_3=\sum_{j=1}^{4} P_{3j}\cdot\log_2\frac{1}{P_{3j}} = H_4=\sum_{j=1}^{4} P_{4j}\cdot\log_2\frac{1}{P_{4j}} = 0.922\ [\text{bits/symbol}]$$

$$H=\sum_{i=1}^{4} P_i\cdot H_i = P(x_1)\cdot H_1+P(x_2)\cdot H_2+P(x_3)\cdot H_3+P(x_4)\cdot H_4 = 124.6\ [\text{Mbits/s}]$$

$$R=H\cdot R_s = 1.246\times10^8 = 1.246\ [\text{Mbits/s}]$$

[2] If you are able to employ some capacity-approaching error-correction coding technique and would like to achieve error-free transmission, what is the minimum channel bandwidth required.

$$C=B\cdot\log_2\left(1+\frac{S_p}{N_p}\right)\ge R$$

$$6B\ge124.6$$

$$B\ge21\text{MHz}$$

[EXAMPLE 4]

A ~~predictable~~ predictive source encoder generates a bit stream at a bit rate of 3.1844 Mb/s, and it is known that the probability of a bit taking the value 0 is $P(0)=p=0.95$. The bit stream is then encoded by a run length encoder (RLC) with a coded length of $n=5$ bits.

[1] Determine the compression ratio of the RLC, and the bit rate after the RLC

Average codeword length d before RLC with $N=2^n-1$

$$d=\sum_{l=0}^{N-1}(l+1)\cdot p^l\cdot(1-p) + Np^N = \frac{1-p^N}{1-p}$$

$$\text{Compress ratio} = \frac{d}{n} = \frac{1-p^N}{n(1-p)} = \frac{1-0.95^{2^5-1}}{5(1-0.95)} = 3.184$$

The bit rate after the RLC

$$R_{RLC} = \frac{3.1844}{3.184} = 1\ [\text{Mbits/s}]$$

[2] Find the encoder input patterns that produce the following encoder output codewords.

11111	31x'0'+'1'
11110	30x'0'+'1'
11101	29x'0'+'1'
11100	28x'0'+'1'
11011	27x'0'+'1'
00001	"01"
00000	'1'

[3] What is the encoder input sequence of the RLC coded Signal 11011|00000|11110 ?

27x'0'+"11"+30x'0'+'1'

~~1~~ 2x'1'