System Identification & Information Theory

ELEC 3004: **Systems**: Signals & Controls

Dr. Surya Singh

Lecture 23

(with material from D. MacKay, B. Lathi, K. Harris, D. Corrigan, and more!)

elec3004@itee.uq.edu.au

http://robotics.itee.uq.edu.au/~elec3004/

May 30, 2016

http://elec3004.com

Lecture Schedule:

Week	Date	Lecture Title
1	29-Feb	Introduction
	3-Mar	Systems Overview
2	7-Mar	Systems as Maps & Signals as Vectors
	10-Mar	Data Acquisition & Sampling
3	14-Mar	Sampling Theory
	17-Mar	Antialiasing Filters
4	21-Mar	Discrete System Analysis
	24-Mar	Convolution Review
	28-Mar	Holiday
	31-Mar	
5	4-Apr	Frequency Response & Filter Analysis
	7-Apr	Filters
6	11-Apr	Digital Filters
	14-Apr	Digital Filters
7	18-Apr	Digital Windows
	21-Apr	FFT
8	25-Apr	Holiday
	28-Apr	Introduction to Feedback Control
9	3-May	Holiday
	5-May	Feedback Control & Regulation
10	9-May	Servoregulation/PID
	12-May	Introduction to (Digital) Control
11	16-May	Digital Control Design & State-Space
	19-May	Observability, Controllability & Stability of Digital Systems
12	23-May	Digital Control Systems: Shaping the Dynamic Response & Estimation
	26-May	Applications in Industry
13	30-May	System Identification & Information Theory
	2-Jun	Summary and Course Review

Announcements

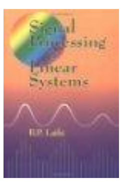
PS 3 Peer Review Competition

- The PS 3 Review with the highest Likert Score
- Deadline for reviews: June 3 (11:59 pm)
- Good reviews discussed June 2nd Last Lecture
- Reward: 3004¢

ELEC 3004 "Review Lab" ("Lab 5"):

- Redo any aspects of any of the labs
- Review course
- Review 2015 Final exam (which we will do on June 2nd also)

Follow Along Reading:



B. P. Lathi
Signal processing and linear systems
1998
TK5102.9.L38 1998

David J. C. MacKay
Information Theory, Inference and Learning Algorithms
2003

http://www.inference.phy.cam.ac.uk/itila/

Today

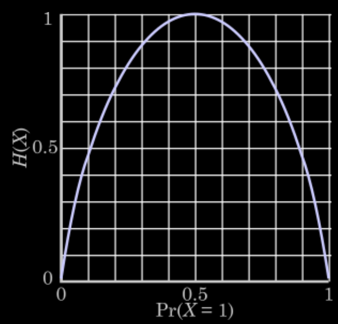
→ State-space ←

- Lathi Ch. 2 (?)
  - § 2.7-6 Time Constant and Rate of Information Transmission
- Information Theory!

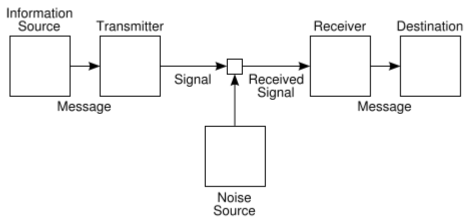
- Final Exam 2015: http://robotics.itee.uq.edu.au/~elec3004/tutes.html#Final

Next Time

Information Theory!



Shannon Information Theory



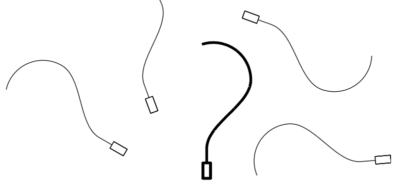
"The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point."

On the transmission of information over a noisy channel:

- An **information source** that produces a message
- A **transmitter** that operates on the message to create a signal which can be sent through a channel
- A **channel**, which is the medium over which the signal, carrying the information that composes the message, is sent
- A **receiver**, which transforms the signal back into the message intended for delivery
- A **destination**, which can be a person or a machine, for whom or which the message is intended

Information theory is…

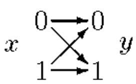
- It all starts with **Probability Theory**!



1. (Wikipedia) `Information theory is a branch of applied mathematics, electrical engineering, and computer science involving the quantification of information.`

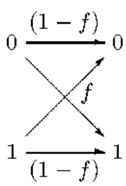
2. Information theory is **probability theory** where you take logs to base 2.

Binary Systematic Channel



$$\begin{aligned} P(y=0 \,|\, x=0) &= 1-f; & P(y=0 \,|\, x=1) &= f; \\ P(y=1 \,|\, x=0) &= f; & P(y=1 \,|\, x=1) &= 1-f. \end{aligned}$$

- Ex: A File of N=10,000 bits is transmitted with ***f*=0.1**

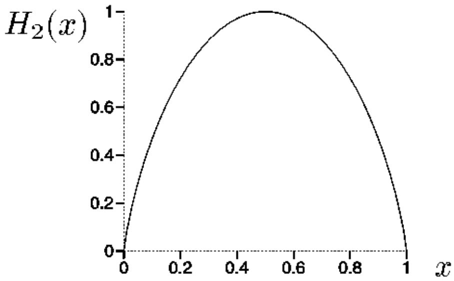


How many bits are flipped?

Binary Systematic Channel

Assume a Binomial distribution

- Mean: $\mu = Np$
- Variance: $\sigma^2 = Npq$



Then:

$$\mu = (0.1)10{,}000$$
$$\sigma^2 = (0.1)(0.9)10{,}000 = 900$$

Thus:

$$1000 \pm 30$$

How to beat this?

Filter like mad!

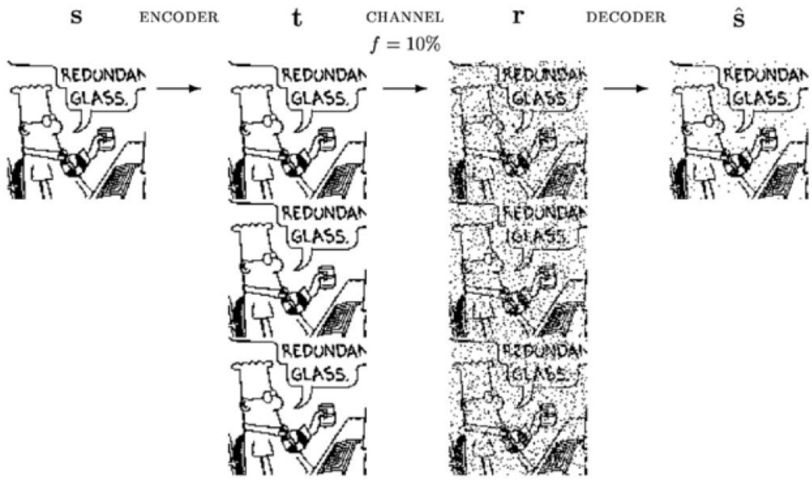
- But this will also affect "non-noisy" signal portions

Matlab: `medfilt2(noisy_im)`

Redundant Transmission!

- Repetition Code 'R3'
  $r = t \boxed{modulo\ 2} n$

Repetition Code 'R3'



Source: MacKay VideoLectures 01, Slide 30

Viewing this as a Likelihood

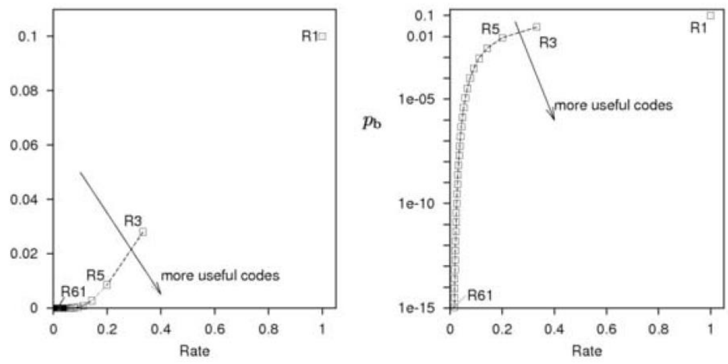
$$\frac{\overbrace{P(r|s)}^{Likelihood\ of\ s}\ \overbrace{P(s)}^{Prior\ Probability\ of\ s}}{\underbrace{P(r)}_{Sum\ of\ P(r)}}$$

- $P(r) = P(r|s=0)P(s=0) + P(r|s=1)P(s=1)$

- $P(r|s=0) = (1-f) \times f \times f$
- $P(r|s=0) = (f) \times (1-f) \times (1-f)$
- $P(s=0) = \frac{1}{2}, P(s=1) = \frac{1}{2}$

➔ $P(s=1|r=011) = \frac{(1-f)^2 f \cdot \frac{1}{2}}{(1-f)f^2 + f(1-f)^2 \cdot \frac{1}{2}} = (1-f)$

Performance of repetition codes



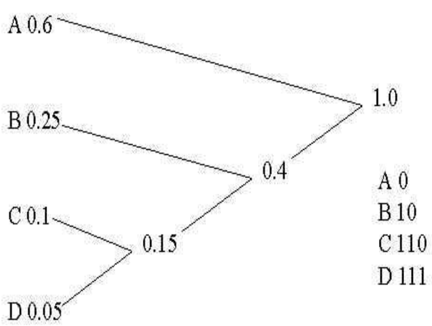
Source: MacKay VideoLectures 01, Slide 45

Huffman Coding

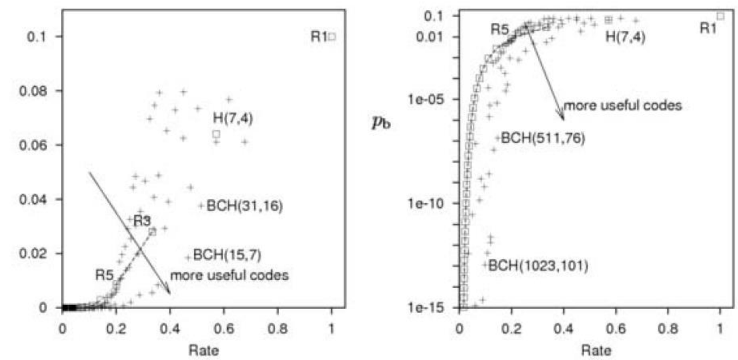
- Huffman is the simplest entropy coding scheme
  - It achieves average code lengths no more than 1 bit/symbol of the entropy

- A binary tree is built by combining the two symbols with lowest probability into a dummy node

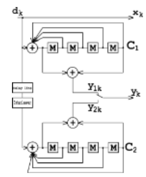
- The code length for each symbol is the number of branches between the root and respective leaf



What's Achievable?

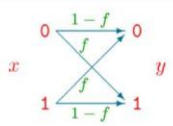


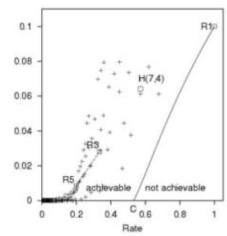
- [Turbocodes](#): Interative [forward error correction](#) (FEC) codes



Source: MacKay VideoLectures 01, Slide 54

Shannon's noisy-channel coding theorem





For any channel:
Reliable (virtually error-free) communication is possible
at rates up to C

$$C_{\mathrm{BSC}}(f) = 1 - H_2(f)$$

$$H_2(f) = f \log_2 \frac{1}{f} + (1-f)\log_2 \frac{1}{1-f}$$

Source: MacKay VideoLectures 01, Slide 58

Entropy

- **Entropy: Probability in Disguise!**

The entropy of a random variable X with a probability mass function $p(x)$ is defined by

$$H(X) = -\sum_{x} p(x)\log_2 p(x). \tag{1.1}$$

Example 1.1.2 Suppose that we have a horse race with eight horses taking part. Assume that the probabilities of winning for the eight horses are $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\right)$. We can calculate the entropy of the horse race as

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{16}\log\frac{1}{16} - 4\frac{1}{64}\log\frac{1}{64}$$
$$= 2 \text{ bits}. \tag{1.3}$$

Entropy

- A message $X_1 \ldots X_n$ has length

$$\sum_i -\log_2 p(X = X_i)$$

- A long message, has an average length per codeword of:

$$H(X) = E(-\log_2 p(X)) = \sum_X -p(X)\log_2 p(X)$$

Entropy is always positive, since $p(X) \geq 0$.

Connection to Physics

- A macrostate is a description of a system by large-scale quantities such as pressure, temperature, volume.
- A macrostate could correspond to many different microstates i, with probability p_i.
- Entropy of a macrostate is
- $S = -k_B \sum_i p_i \ln p_i$
- Hydrolysis of 1 ATP molecule at body temperature: ~ 20 bits

Entropy: So why does this work?

Example 2.1.1 Let

$$X = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases} \tag{2.4}$$

Then

$$H(X) = -p \log p - (1-p)\log(1-p) \stackrel{\text{def}}{=\!=} H(p). \tag{2.5}$$

Plot a graph of H(p) against p.

Entropy

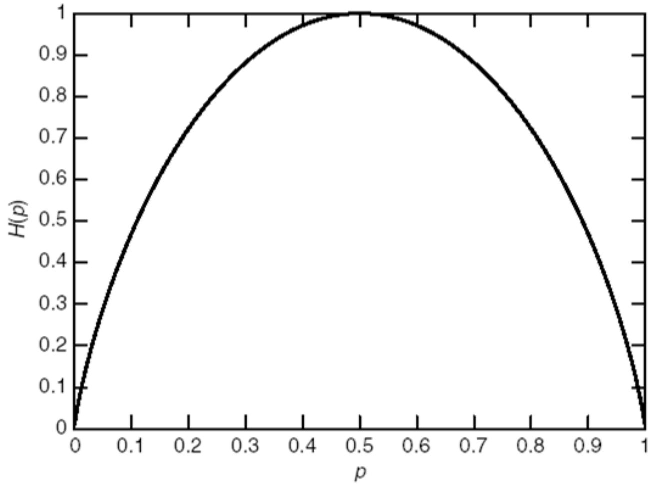


FIGURE 2.1. $H(p)$ vs. p.

Maximum entropy distributions

- Entropy measures uncertainty in a variable.
- If all we know about a variable is some statistics, we can find a distribution matching them that has maximum entropy.
- For constraints $E(S_i) = s_i, i = 1 \dots n$, usually of form

$$p(X) = e^{-\sum_i \lambda_i s_i}$$

- Relationship between λ_i and s_i not always simple
- For continuous distributions there is a (usually ignored) dependence on a reference density – depends on coordinate system

Examples of maximum entropy distributions

Data type	Statistic	Distribution
Continuous	Mean and variance	Gaussian
Non-negative continuous	Mean	Exponential
Continuous	Mean	Undefined
Angular	Circular mean and vector strength	Von Mises
Non-negative Integer	Mean	Geometric
Continuous stationary process	Autocovariance function	Gaussian process
Point process	Firing rate	Poisson process

Conditional Entropy

- Suppose Alice wants to tell Bob the value of X
  - And they both know the value of a second variable Y.
- Now the optimal code depends on the conditional distribution $p(X|Y)$
- Code length for $X = i$ has length $-\log_2 p(X = i|Y)$
- Conditional entropy measures average code length when they know Y

$$H(X|Y) = -\sum_{X,Y} p(X,Y) \log_2 p(X|Y)$$

Mutual information

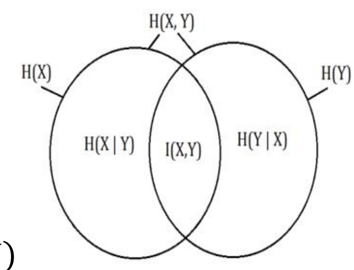
- How many bits do Alice and Bob save when they both know Y?

$$I(X;Y) = H(X) - \mathrm{H(X|Y)}$$
$$= \sum_{X,Y} p(X,Y)(-\log_2 p(X) + \log_2 p(X|Y))$$
$$= \sum_{X,Y} p(X,Y) \log_2 \left(\frac{p(X,Y)}{p(X)p(Y)}\right)$$

- Symmetrical in X and Y!
- Amount saved in transmitting X if you know Y equals amount saved transmitting Y if you know X.

Properties of mutual information

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



- If $X = Y$, $I(X;Y) = H(X) = H(Y)$

- If X and Y are independent, $I(X;Y) = 0$

$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$ $H(X|Y) \leq H(X)$

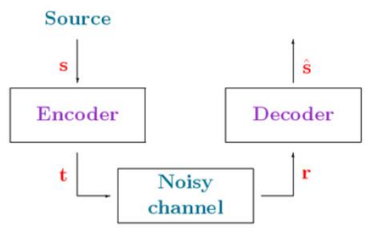
Data Compression

- How to measure information Content?

Compression:

- Add redundancy

- Do Inference



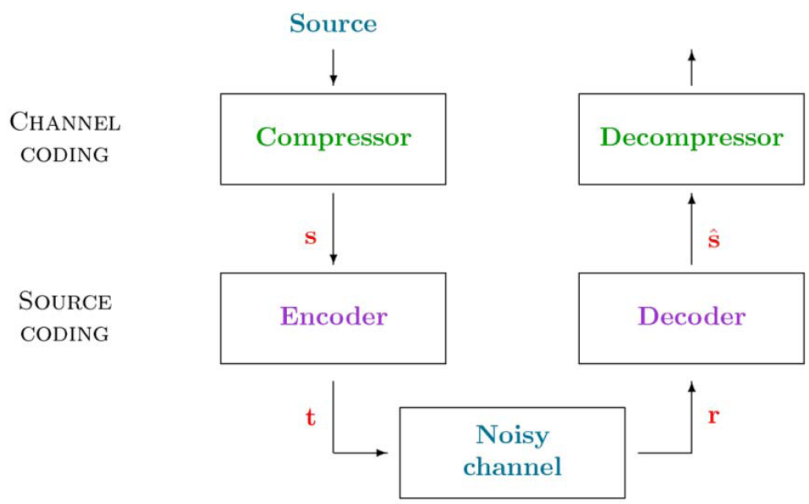
Example: Mystery Text

```
I.    Emma Woodh*use, hands*me, clever* and rich,*with a comiortab*e home an*
      happy di*position,*seemed to*unite som* of the b*st bless*ngs of
      e*istence;*and had *ived nea*ly twenty *ne year* in the*world w*th
      very*little *0 distr*ss or vex*her. *he was*the yo*ngest *f the *wo
      dau*hters *f a most *ffect*onate* indu*gent *ather* and *ad, i* cons*quenc*
      of h*r si*ter'* mar*iage* bee* mis*ress*of h*s ho*se f*om a ver* ea*ly
      *eri*d. *er *oth*r h*d d*ed *00 *ong*ago*for*her to*ha*e *or*
      t*an*an*in*is*in*t *em*mb*an*e *f *er*ca*es*es* a*d*h*r*p*a*e*h*d*b*e*
      *u*p*i*d*b* *n*e*c*l*e*t*w*m*n*a* g**e**e**,**h**h**
      **l**n**i**l**s**r**o**a**o**e**i*
      a***c***n***S***e***y***s***d***s***a***r***e***n***
      W****o****s****i****l****a****g****n****t****a****e****v***

II.   Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy
      disposition, seemed to unite some of the best blessings of existence; and had lived
      nearly twenty one years in the world with very little to distress or Vex her. She was
      the youngest of the two daughters of a most affectionate, indolent father; and had ,
      in consequence of her sister's marriage , been mistress of his house from a very
      early period. Her mother had died too lone ago for her to have more than an
      indistinct remembrance of her caresses; and her place had been supplied by an
      excellent woman as governess, who had fallen little short of a mother in affection.
      Sixteen years had Miss Taylor been in Mr Woodhouse's family , 1ess as a eoverness
      than a friend , very
```

Source: MacKay VideoLectures 02, Slide 5

Channel and Source Coding



Ex: Bent/Unfair Coin

- A simple redundant source - a bent coin…

```
000000000000000001001001000000000000000000001000010001010000000000101000000000000000
000001000000000010000000000000010000011000001000000000000000010000000000010000100001
000000000000000000001000000100000000000010000000000000000000000000010000000000000000
010001000000100000010010000000000001000000001000000000000100100001010100000000110000
100000000000010000000000000000000000000000000000000000101000000100000000000000010
001010000110000000001000000000000000000000000000000000000000000000000010000000000
000100000000000000000001000010000000000110110000000101000101000000000000000000000000
001000101000001000010000000000000000000100000100000100001001000000001010001000000000
000000000000000000000000000001100000000101000000000010010001010100011001010100000000
000000100000000000000000001000000010100001101000000001000000010110000000010000000100
100000100000000000000010000000000000001000001000000000000000100000000000000000000001000
001000100000001000000000000000100000000000100000100000000000001101000001001000000000000
000000001000000000000000000100000000010000000000000000000000001000000010010001000000000
011000000011100000000000000010000000000000000000000000100000000000000001000000000000000000
001000000000000101000000100000000000000100000000000000100010100010010001000000000000000000
000000000000000000000101000000000100100000000000000000000000000100000110000000100000000
010000000000000001000100000000000001000000000000000000000000000000010000000000000000100
000001000000000000000000000000000000000000000001000000000100100000000010000000000
1000010000000100000000000
```

- $P_1 = 0.1$

Ben/Unfair Coin

- Outcome: $x \in \{Tails, Heads\}$
- Probabilities $\{0.9, 0.1\}$

$$P(x = Tails) = 0.9$$

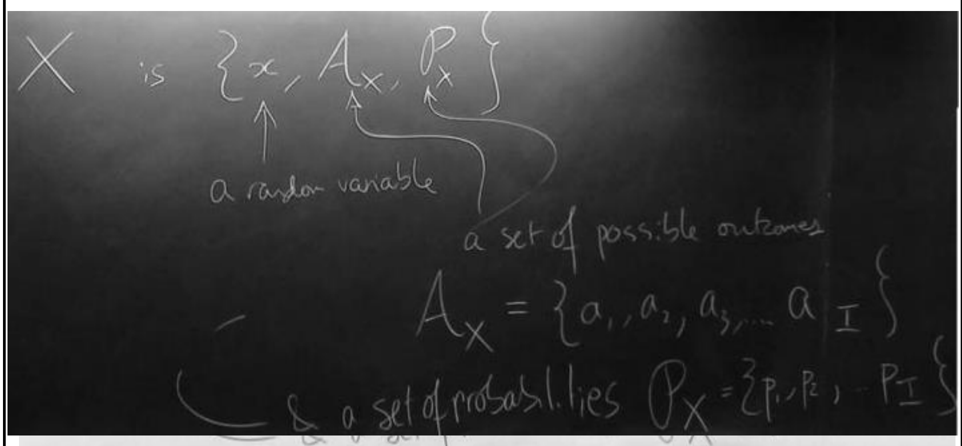
Such that:

- $P(x = a_i) = p_i$

$$\sum_i p_i = 1$$

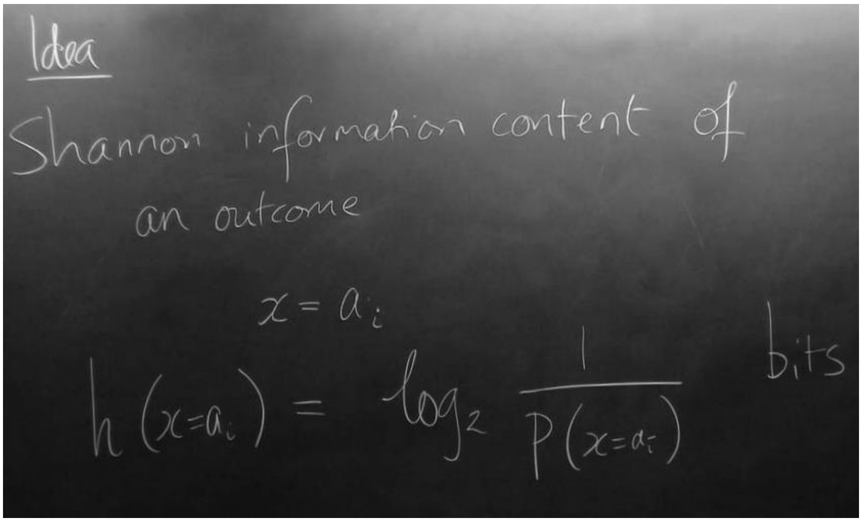
Source: MacKay VideoLectures 02, Slide 10

Ben/Unfair Coin



Source: MacKay VideoLectures 02, Slide 12

Ben/Unfair Coin

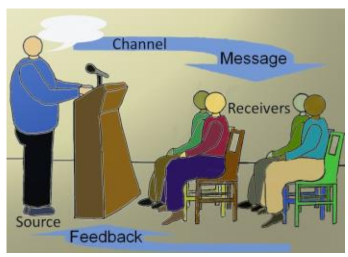


Source: MacKay VideoLectures 02, Slide 14

Shannon and Weaver: Models of Communication

Three "problems" in Communication:

- The technical problem: how accurately can the message be transmitted?
- The semantic problem: how precisely is the meaning "conveyed"?
- The effectiveness problem: how effectively does the received meaning affect behaviour?

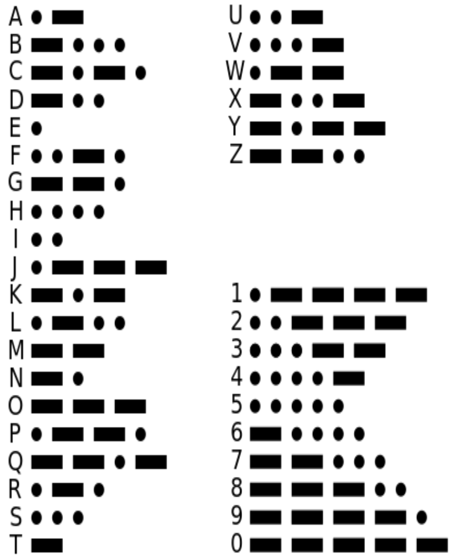


Source: http://en.wikipedia.org/wiki/File:Transactional_comm_model.jpg

Ex: Morse code

- Code words are shortest for the most common letters

- This means that messages are, on average, sent more quickly.



What is the "optimal code"?

- X is a random variable

- Alice wants to tell Bob the value of X (repeatedly)

- What is the best binary code to use?

- How many bits does it take (on average) to transmit the value of X?

Optimal code lengths

- In the optimal code, the word for $X = i$ has length
- $\log_2 \frac{1}{p(X=i)} = -\log_2 p(X = i)$
- For example:

Value of X	Probability	Code word
A	½	0
B	¼	10
C	¼	11

- ABAACACBAACB coded as 010001101110001110
- If code length is not an integer, transmit many letters together

Kullback-Leibler divergence

- Measures the difference between two probability distributions
  - (Mutual information was between two random variables)
- Suppose you use the wrong code for X. How many bits do you waste?

Length of codeword

Length of optimal codeword

- $D_{KL}(p||q) = \sum_x p(x)\left[\log_2 \frac{1}{q(x)} - \log_2 \frac{1}{p(x)}\right]$
$$= \sum_x p(x) \log_2 \frac{p(x)}{q(x)}$$
- $D_{KL}(p||q) \geq 0$, with equality when p and q are the same.
- $I(X;Y) = D_{KL}(p(x,y)||p(x)p(y))$

Continuous variables

- X uniformly distributed between 0 and 1.
- How many bits required to encode X to given accuracy?

Decimal places	Entropy
1	3.3219
2	6.6439
3	9.9658
4	13.2877
5	16.6096
Infinity	Infinity

- Can we make any use of information theory for continuous variables?

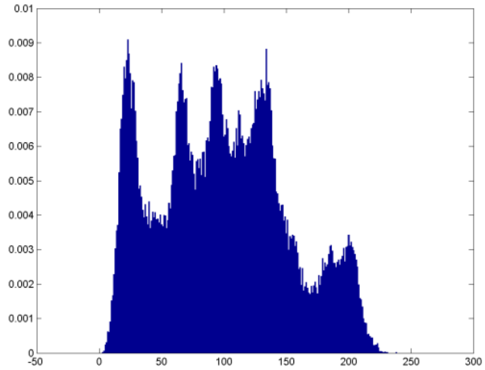
K-L divergence for continuous variables

- Even though entropy is infinite, K-L divergence is usually finite.

- Message lengths using optimal and non-optimal codes both tend to infinity as you have more accuracy. But their difference converges to a fixed number.

$$\sum_x p(x)\log_2\frac{p(x)}{q(x)} \rightarrow \int p(x)\log_2\frac{p(x)}{q(x)}dx$$

Calculating the Entropy of an Image

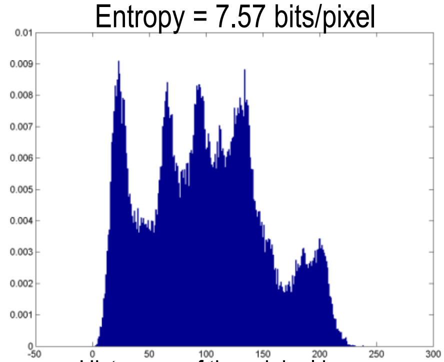


The entropy of lena is = 7.57 bits/pixel approx

Entropy In General

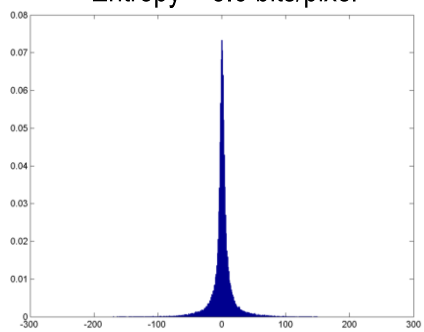
- Entropy of a source is maximised when all signals are equiprobable and is less when a few symbols are much more probable than the others.

Entropy = 7.57 bits/pixel



Histogram of the original image

Entropy = 5.6 bits/pixel



Histogram of the difference image

Huffman Coding of Lenna

Symbol	Code Length
0	42
1	42
2	41
3	17
4	14
…	…

Average Code Word Length = $\sum_{k=0}^{255} p_k l_k = 7.59\ bits/pixel$

So the code length is not much greater than the entropy

But this is not very good

- Why?
  - Entropy is not the minimum average codeword length for a source with memory
  - If the other pixel values are known we can predict the unknown pixel with much greater certainty and hence the effective (ie. conditional) entropy is much less.
- Entropy Rate
  - The minimum average codeword length for any source.
  - It is defined as

$$H(\chi) = \lim_{n\to\infty} \frac{1}{N} H(X_1, X_2, \ldots, X_n)$$

Coding Sources with Memory

- It is very difficult to achieve codeword lengths close to the entropy rate
  - In fact it is difficult to calculate the entropy rate itself
- We looked at LZW as a practical coding algorithm
  - Average codeword length tends to the entropy rate if the file is large enough
  - Efficiency is improved if we use Huffman to encode the output of LZW
  - LZ algorithms used in lossless compression formats (eg. .tiff, .png, .gif, .zip, .gz, .rar…)

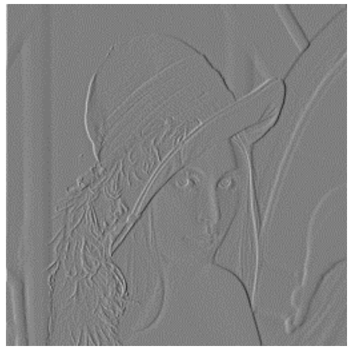
Efficiency of Lossless Compression

- Lenna (256x256) file sizes
  - Uncompressed tiff - 64.2 kB
  - LZW tiff – 69.0 kB
  - Deflate (LZ77 + Huff) – 58 kB

- Green Screen (1920 x 1080) file sizes
  - Uncompressed – 5.93 MB
  - LZW – 4.85 MB
  - Deflate – 3.7 MB

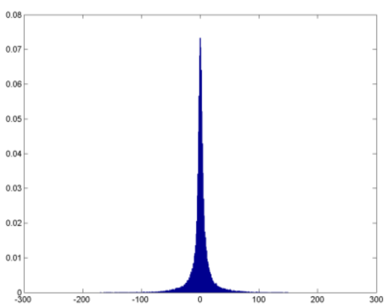
Differential Coding

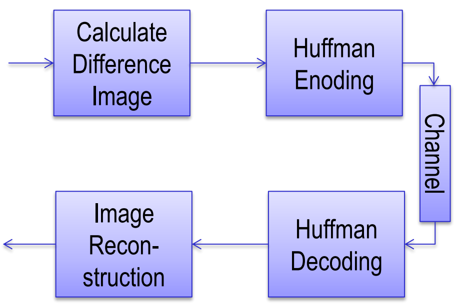
- Key idea – code the differences in intensity.



$G(x,y) = I(x,y) - I(x-1,y)$

Differential Coding





- The entropy is now 5.60 bits/pixel which is much less than 7.57 bits/pixel we had before (despite having twice as many symbols)

Lossy Compression

- But this is still not enough compression
  - Trick is to throw away data that has the least perceptual significance

Effective bit rate = 8 bits/pixel

Effective bit rate = 1 bit/pixel (approx)

Next Time…

- **Information Theory & More!**

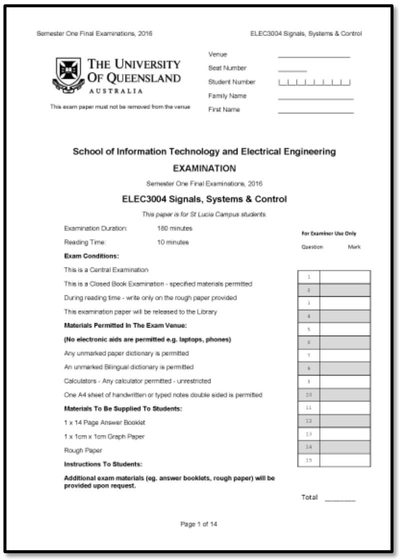
- Review:
  - Chapter 6 of FPW
  - Chapter 13 of Lathi

- Deeper Pondering??

Final Exam Review

- June 10, 2016
- From: 2-4
- In: 45-204 (???)
- [Some Review Notes (from Course Textbooks)](http://robotics.itee.uq.edu.au/~elec3004/tutes.html)

→http://robotics.itee.uq.edu.au/~elec3004/tutes.html



Semester One Final Examinations, 2016 — ELEC3004 Signals, Systems & Control

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

This exam paper must not be removed from the venue

Venue ______

Seat Number ______

Student Number ______

Family Name ______

First Name ______

School of Information Technology and Electrical Engineering

EXAMINATION

Semester One Final Examinations, 2016

ELEC3004 Signals, Systems & Control

This paper is for St Lucia Campus students.

Examination Duration: 180 minutes

Reading Time: 10 minutes

Exam Conditions:

This is a Central Examination

This is a Closed Book Examination - specified materials permitted

During reading time - write only on the rough paper provided

This examination paper will be released to the Library

Materials Permitted In The Exam Venue:

(No electronic aids are permitted e.g. laptops, phones)

Any unmarked paper dictionary is permitted

An unmarked Bilingual dictionary is permitted

Calculators - Any calculator permitted - unrestricted

One A4 sheet of handwritten or typed notes double sided is permitted

Materials To Be Supplied To Students:

1 x 14 Page Answer Booklet

1 x 1cm x 1cm Graph Paper

Rough Paper

Instructions To Students:

Additional exam materials (eg. answer booklets, rough paper) will be provided upon request.

For Examiner Use Only

Question	Mark
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

Total ______

Page 1 of 14