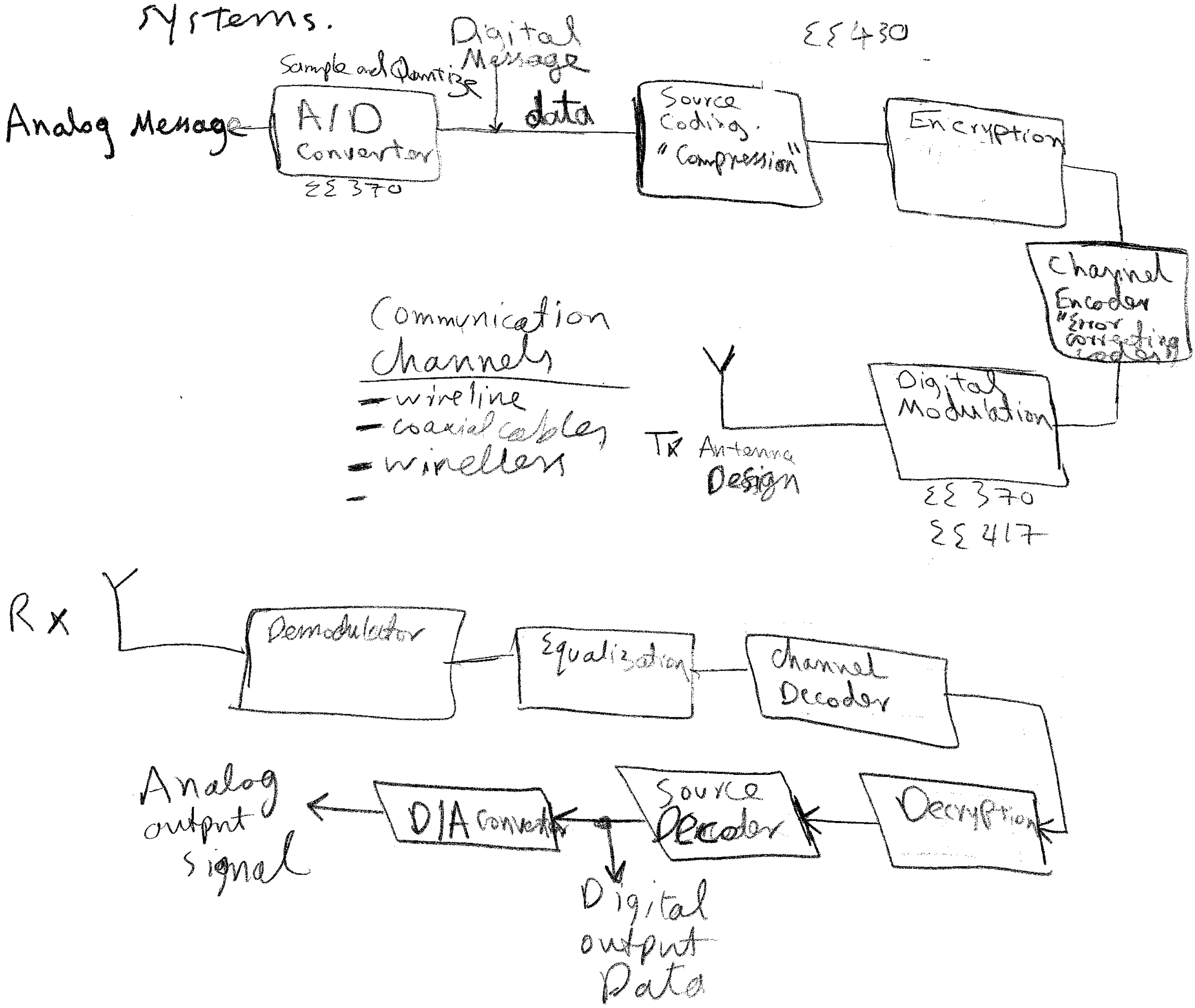
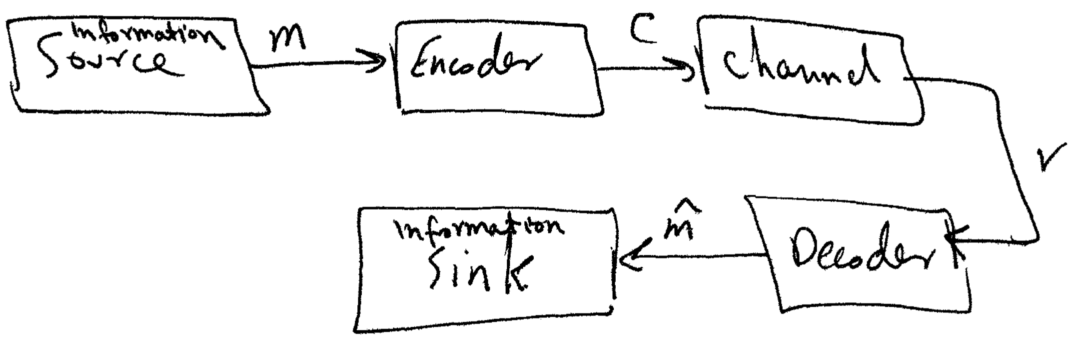
Introduction Class

- Let take a quick Review A Digital Communication systems.



Chapter 1

Discrete Sources and Entropy



1.2 Source Alphabets and Entropy.

Def. — An information Source outputs a set of symbols.

— A finite discrete source ~~is one~~ output ~~is one~~ belongs to a finite discrete set of symbols. The symbol set is called the source alphabet.

— $A = \{a_0, a_1, \ldots, a_{M-1}\}$ is a source alphabet (Symbol set) of cardinality $M = |A|$. (size) or (number of elements)

— The source outputs symbols in a time sequence represented by the notation $\bar{a} = (s_0 s_1 \cdots s_t \cdots)$ where $s_t \in A$

— The probability that the source emits symbol a_m is written as $P_m = Pr(a_m)$.

— The set of probabilities for the source alphabet is

$$P_A = \{P_0, P_1, \cdots, P_{M-1}\}$$

— Information Theory makes an important distinction between data and information. Not all data carries information. Information is the part that adds ~~to~~ knowledge.

~~Given a set of data~~

Example

Assume that we have a source that emits only one symbol. ~~and with~~ A set of data from this source will have zero information content since the receiver knows that the source only send one unique symbol all the time. Thus, we have a set of data but no information content.

— From the above example, we can see that the information content of the data increases with uncertainty of transmitted symbols. —

— Entropy

Information theory provides a measure of the average amount of information conveyed per source symbol. This measure is called the entropy of the source and is defined as:

$$H(A) = \sum_{m=0}^{M-1} P_m \log_2\left(\frac{1}{P_m}\right)$$

; where P_m is the probability that symbol m was transmitted. Notice that the Logarithm is to the base 2.

the units of $H(A)$ is bits. It tells us How much information carried by each symbol.

To help in calculation, Recall that $\log_2(x) = \frac{\ln(x)}{\ln(2)}$

Also, $\lim_{x \to 0} x \log(x) = 0$

Notice that the information per symbol is $I(A) = \log_2\left(\frac{1}{P_m}\right)$ and Entropy is the average information per

Example 1.2.1

What is the average amount of information conveyed per source symbol of a 4-ary source having probabilities

$$P_A = \{0.5, 0.3, 0.15, 0.05\} \;?$$

Solution:

$$H(A) = 0.5\log_2(2) + 0.3\log_2(10/3) + 0.15\log_2\left(\frac{100}{15}\right) + 0.05\log_2(20)$$

$$= 1.6477 \text{ bits.}$$

So, each symbol carries 1.6477 bits of information. while it carries 2 bits of data.

Therefore, it is possible to use some data compression techniques to compress the above source and use fewer bits on average.

Def: Information efficiency $= \dfrac{\text{The entropy of the source}}{\text{Average number of bits used to represent the source data}}$.

For the above example,

the information efficiency is $= \dfrac{1.6477}{2} = 82.387\%$

this means that approximately 17.6% of the bits are redundant and carry no information.

Example; Find the entropy of the 4-ary signal if all symbols are equally probable. i.e $P_m = \frac{1}{4}$

Lemma: Consider an M-ary source A, the maximum entropy of this source is $\log_2 M$ and it happens when all symbols are equally probable $\Rightarrow P_m = \frac{1}{M}$ for all $m \in A$

See example 1.2.2 for proof.

This result makes sense intuitively. If every symbol in A is equally probable, an observer would have no idea what symbol will be emitted next by the source. Thus, each symbol carries the maximum surprise value and the average amount of information is maximized.

1.2.2 Joint and Conditional Entropy

Most communication systems are designed to be used by a large number of users. The designers of such a system are concerned with maximizing the total information carrying capacity of the system.

Information theory gives us also tools to measure the joint and conditional entropy for more than one source.

Consider a situation where we have two information sources, A and B. Let $|A| = M_A$ and $|B| = M_B$.

The joint probability that A sends symbol a_i and B sends symbol b_j is $P_{ij} = Pr(a_i, b_j)$

If the two symbol are statistically independent, then

Statistically independent case

If sources A and B are statistically independent, the total entropy of this system will be

$$H(A,B) = H(A) + H(B)$$

joint entropy Entropy for A Entropy for B

If the two sources are dependent, what do you expect will happen to the joint entropy? will it increase or decrease?
The answer is that it will decrease. There is a fundamental theorem in information theory that says " Side information never increases entropy". The joint entropy

$$H(A,B) \leqslant H(A) + H(B)$$

with equality if and only if A and B are statistically independent.

Statisically dependent Case

Denote the combined emission of a_i and b_j as a compound symbol $c_{i,j} \equiv \langle a_i, b_j \rangle$. let the probability of emiting $c_{ij} = P_{ij}$.

The entropy of C is

~~joint~~ joint entropy $\longrightarrow$ $$H(C) = \sum_{c_{ij} \in C} P_{ij} \log_2\left(\frac{1}{P_{ij}}\right)$$

$$\equiv \sum_{i=0}^{M_A-1} \sum_{j=0}^{M_B-1} P_{ij} \log_2\left(\frac{1}{P_{ij}}\right)$$

The joint probability P_{ij} may be written in terms of a conditional probability $\left[P_{j|i} = \Pr(b_j|a_i)\right]$

as $$P_{ij} \equiv P_{j|i} \cdot P_i$$

$$\therefore\ H(C) = \sum_i \sum_j P_{ij} \log_2\left(\frac{1}{P_{j|i} P_i}\right)$$

$$= \sum_i \sum_j P_{ij} \log_2\left(\frac{1}{P_i}\right) + \sum_i \sum_j P_{ij} \log_2\left(\frac{1}{P_{j|i}}\right)$$

Notice that, $$\sum_i \sum_j P_{ij} \log_2\left(\frac{1}{P_i}\right) = \sum_i \sum_j P_{j|i} P_i \log_2\left(\frac{1}{P_i}\right)$$

$$= \sum_i P_i \log_2\left(\frac{1}{P_i}\right) \underbrace{\sum_j P_{j|i}}_{=1}$$

$$\therefore\ H(C) = \sum_i P_i \log_2\left(\frac{1}{P_i}\right) + \sum_i \sum_j P_{ij} \log_2\left(\frac{1}{P_{j|i}}\right)$$

$$H(C) = H(A,B) = H(A) + H(B|A)$$

Similarly, $$H(A,B) = H(B) + H(A|B)$$

H(B/A) and H(A/B) are conditional Entropy.

$$H(B/A) = \sum_i \sum_j P_{ij} \log\left(\frac{1}{P_{j|i}}\right)$$

$$H(A/B) = \sum_i \sum_j P_{ij} \log\left(\frac{1}{P_{i|j}}\right)$$

Also, $H(B/A) < H(B)$

and $H(A/B) < H(A)$

The end result is that the joint entropy of two sources A and B is

$$H(A,B) = H(A) + H(B/A)$$
$$= H(B) + H(A/B)$$

and

$$H(A,B) \le H(A) + H(B)$$

with equality if and only if A and B are independent.

Continue 1.2.2

— Joint and conditional Entropy.

$$H(C) = H(A,B) = H(A) + H(B|A)$$

joint Entropy. conditional Entropy.

$$= H(B) + H(A|B)$$

$$H(A,B) \leq H(A) + H(B)$$

and equality happens when A and B are independent.

Example 1.2.3

Error detection using parity bits

parity bits are added to the transmitted bits to detect errors. Let A be an information source with alphabet $A = \{0, 1, 2, 3\}$. Let each symbol a be equally probable and Let $B = \{0, 1\}$ be a parity generator with

$$b_j = \begin{cases} 0 & \text{if } a = 0 \text{ or } a = 3 \\ 1 & \text{if } a = 1 \text{ or } a = 2 \end{cases}$$

What are $H(A)$, $H(B)$ and $H(A,B)$?

$$H(A) = 4 \cdot \tfrac{1}{4} \log_2(4) = 2 \text{ bit}$$

Likewise $H(B) = 2 \cdot \frac{1}{2} \log_2(2) = 1$ bit

This is because each symbol is equally probable.

The conditional probabilities $Pr(b/a)$ are

$Pr(0|0) = 1$, $Pr(1/0) = 0 \rightarrow \sum_j Pr(b_j/0) = 1$

$Pr(0|1) = 0$, $Pr(1/1) = 1$

$Pr(0/2) = 0$, $Pr(1/2) = 1$

$Pr(0/3) = 1$, $Pr(1/3) = 0$

Therefore, $$H(B/A) = \sum_{i=0}^{3} P_i \sum_{j=0}^{1} P_{j|i} \log_2\left(\frac{1}{P_{j|i}}\right)$$

$$= 4 \cdot \tfrac{1}{4} (1 \log_2 1 - 0 \log_2 0)$$

This says that B is completely determined by A.

$$H(A,B) = H(A) + H(B|A) = 2+0 = 2$$

$\therefore$ Source B contributes no information to the compound signal.

1.7.3 Entropy of Symbol Blocks and the Chain Rule.

What is the information content of a block of symbols?

A block of symbols is a sequence of n symbols $(s_0, s_1, \dots, s_{n-1})$ produced by source A. $\{s_t \in A\}$.

The entropy of this block is denoted as

Joint Entropy $\rightarrow$ $H(A_0, A_1, \dots, A_{n-1})$.

We can use the joint entropy result to express the entropy as

$$H(A_0, A_1, \dots, A_{n-1}) = H(A_0) + H(A_1, A_2, \dots, A_{n-1}|A_0)$$

This term can also be written as $\rightarrow$ $H(A_1, A_2, \dots, A_{n-1}|A_0) = H(A_1|A_0) + H(A_2, \dots, A_{n-1}|A_0, A_1)$

Repeating this argument inductively, we get

$$H(A_0, A_1, \dots, A_{n-1}) = H(A_0) + H(A_1|A_0) + H(A_2|A_0, A_1) + \dots H(A_{n-1}|A_0, \dots, A_{n-2})$$

This is called the chain rule for entropy..

Since $H(B|A) \leq H(B)$

The upper bound on the Entropy of symbol blocks is

$$H(A_0, \cdots, A_{n-1}) \leq \sum_{i=0}^{n-1} H(A_i)$$

with equality if and only if all ~~of~~ the symbols in the equence are <u>statistically independent</u>.

<u>Memoryless Source</u>

A source that emits statistically independent symbols is called a memoryless source. That is because it doesn't remember from one time to the next what symbols it has previously emitted. So, there is no correlation between symbols.

If the information source is the same ~~for~~ all the time and the probabilities of the source don't change over time, then we have the joint entropy to be.

$$H(A_0, \cdots, A_{n-1}) \leq n \cdot H(A)$$

~~See~~ Examples 1.2.4 :

Suppose a memoryless source with $A = \{0, 1\}$ having equal symbol probabilities emits a sequence of six symbols. Following the sixth symbol, a seven symbol is added which is ~~the~~ sum modulo 2 of the six previous symbols. What is the entropy of this sequence?

Solution

Let $b = \sum_{t=0}^{5} \oplus s_t$, where $\sum \oplus$ is just the summation Modulo 2 and $s_t \in A$.

$$H(A_0, A_1, \ldots, A_5, b) = H(A_0) + H(A_1|A_0) + \cdots + H(b|A_0, \ldots, A_5)$$

since the First six symbols are statistically independent, we ~~have~~ get

$$H(A_0, A_1, \ldots, A_5, b) = 6\,H(A) + H(b|A_0, \ldots, A_5)$$

since b is completly determined by b, it has no new information and $H(b/A_0, \ldots, A_5) = 0$

$\Rightarrow$ The ~~joint~~ entropy of the ~~s~~block is

$$H(A_0, A_1, \ldots, A_5, b) = 6\,H(A)$$
$$= 6$$

since the symbols of A are equally probable.