ENTROPY CODING

The design of a variable-length code such that its average code word length approaches the entropy of the DMS is often referred to as enlmpy coding. In this section we present two examples of entropy coding.

- Shannon- Fano Coiding
- Huffman Coding

$\boxed{\text{A. Shannon-Fun Coding:}}$ An efficient code can be obtained by the following simple procedure, known as Shannon- Fano algorithm:

1. List the source symbols in order of decreasing probability.
2. Partition the set into two sets that are as close to equiprobable as possible, and assign 0 to the upper set and 1 to the lower set.

B. Huffman Encoding:

In general, Huffman encoding results in an optimum code. Thus, it is the code that has the highest efliciency.

The Huffman encoding procedure is as follows:

1. List the source symbols in order of decreasing probability.
2. Combine the probabilities of the two symbols having the lowest probabilities, and reorder the resultant probabilities; this step is called reduction 1. The same procedure is repeated until there are two ordered probabilities remaining.
3. Start encoding with the last reduction, which consists of exactly two ordered probabilities. Assign 0 as the first digit in the code words for all the source symbols associated with the first probability; assign 1 to the second probability.
4. Now go back and assign 0 and 1 to the second digit for the two probabilities that were combined in the previous reduction step, retaining all assignments made in Step 3.
5. Keep regressing this way until the first column is reached.

10.5. A high-resolution blackand-white TV picture consists of about 2×10^6 picture elements and 16 different brightness levels. Pictures are repeated at a rate of 32 per second. All picture elements are assumed to be independent, and all levels have equal likelihood of occurrence. Calculate the average rate of information conveyed by this TV picture source.

Channel Capacity

A. Channel Capacity per Symbol C:
The channel cizpucily per symbol of a DMC is detined as

$$C. = \max I(X;Y) \qquad \text{b/symbol}$$

where the maxiniization is over all possible input probability distributions P(x,) on X. Note that the channel capacity CA is a function of only the channel transition probabilities that define the channel.

Channel Capacity

B. Channel Capacity per Second C:

If r symbols are being transmitted per second, then the maximum rate of transmission of information per second is rC¿.. This is the channel capacity per secvml and is denoted by C (bls): C : rC, b/s (10.34)

C. Capacities of Special Channels:

1. Lossless Channel:

For a lossless channel, H(X—Y) Z 0 $I(X;Y):11(X)$ Thus, the mutual information (information transfer) is equal to the input (source) entropy, and no source infomation is lost in transmission. Consequently. the channel capacity per symbol is

$$Cs : maxH(X) = logzm$$

where m is the number of symbols in X.

2. Deterministic Channel: For a deterministic channel, H(Y—X) = 0 for all input distributions Pm), and I(X; Y) : H(Y) (10.37) Thus. the information transfer is equal to the output cntropy. The channel capacity per symbol is C. : max H(Y) = 1og2 yi (10.38) iron) where ri is the number of symbols in Y.

3. Naireless Channel:
Since a noiseless channel is both lossless and deterministic, we have I(X; Y) : H(X) = H(Y) (l0.39) and the channel capacity per symbol is C, : logzm :1ogZ n (10440) **4. Binary Symmetric Channel:**
For the BSC o1‘Fig. 10-5. the mutual infomation is (Prob. 10.16) 1(X: Y) ; 11(Y)-l-p1g;1+(11¿)l¡¿2;;(lp) (1041) and the channel capacity per symbol is Ci = 1 +1¿z21+(l p)l¡¿g2(1 p) (10.42)

The unit of I(x)) is the bit (binary unit) if b Z 2, Hartley or decit if b = 10, and nat (natural unir) if b Z at It is standard to use b Z 2. Here the unit bit (abbreviated ”b) is a measure ofinlbrrnation content and is not to bc confused with the term hit meaning ”binary digit.” The conversion of these units to other units can be achieved by the following relationships.

Average Information or Entropy

- In a practical communication system, we usually transmit long sequences of symbols from an information source.
- Thus, we are more interested in the average infornation that a source produces than the information content of a single symbol.
- The mean value of $l(x_i)$ over the alphabet of source X with tn different symbols is given by

Information Rate

If the time rate at which source X emits symbols is r (symbols/s), the Information rate R of the source is given by

MUTUAL INFORMATION

A. Conditional and Joint Entropies:
Using the input probabilities P(x,), output probabilities $P(y_i)$, transition probabilities P(yJ—¿r,), and joint probabilities P(x,, yy), we can define the following various entropy functions for a channel with m inputs and n outputs:

- H(X) = - X P(xi) log
- H(Y) = -2P(yj)]
- H¡X I Y) = - X X
- H = -2 ZP
- H(X, Y) - -2 Z F

Conditional and Joint Entropy

These entropies can be interpreted as follows: H(X) is the average uncertainty of the channel input, and H(Y) is the average uncertainty of the channel output. The conditional entropy H(X]Y) is a measure of the average uncertainty remaining about the channel input after the channel output has been observed. And H(X] Y) is sometimes called the equivncation of X with respect to Y.

- The conditional entropy H(Y—X) is the average uncertainty of the channel output given that X was transmitted.
- The joint entropy H(X, Y) is the average uncertainty of the communication channel as a whole.

Two useful relationships among the above various entropies are

- $H(X,Y) = H(X|Y) + H(Y)(10,26)$ $H(X,Y) = H(Y|X) + H(X)(10.27)$

B. Mutual Information: The mutual information 1(X; Y) of a channel is deiined by I(X; Y) = H(X) H(X—Y) b/symbol (10.28)

Self Information

Self-information This is defined by the following mathematical formula:$I(A) = logbP(A)$
The self-information of an event measures the amount of ones surprise evoked by the event. The negative logarithm $logbP(A)$ can also be written as

$$log_b \frac{1}{P(A)}$$

Note that log(1) = 0, and that $|log(P(A))|$ increases as P(A) decreases from 1 to 0. This supports our intuition from daily experience. For example, a low-probability event tends to cause more “surprise”.

Code efficiency and Code redundancy

The parameter L represents the average number of bits per source symbol used in the source coding process. The code efficiency is defined as

$$\nu = \frac{L_{min}}{L}$$

where L_{min} is the minimum possiblve value of L. When ν approaches unity, the codes is said to be efficient. The code redundancy γ is defined as $\gamma = 1 - \nu$.

Source Coding Theorem

The source coding theorem states that for zi DMS X with entropy $H(X)$, the average code word length L per symbol is bounded as L 2 H(X) (10.52) and further, L can bc made as close to H(X) as dcsircd for some suitably chosen code. Thus, with$L_{min} \geq H(X)$.
The code efficiency can be rewritten as

$$\nu = \frac{H(X)}{L}$$

Kraft inequality

- Let X be a DMS with alphabet ($x_i = \{1, 2, \ldots, m\}$). Assume that the length of the assigned binary code word corresponding to x, is n,.
- A necessary and sufficient condition for the existence of an instantaneous binary code is

$$K = \sum_{i=1}^{m} 2^{-n_i} \leq 1$$

which is known as the **Kraft inequality**.

- Note that the Kraft inequality assures us of the existence of an instantaneously decodable code with code word lengths that satisfy the inequality. But it does not show us how to obtain these code words, nor does it say that any code that satisfies the inequality is automatically uniquely decodable