Shannon-Fano Coding

Presentation by:
Joe Louis Paul I

Shannon-Fano Coding

- Shannon-Fano encoding is the first established and widely used encoding method.
- This method and the corresponding code were invented simultaneously and independently of each other by C. Shannon and R. Fano in 1948.
- The simplest variable length coding algorithm.
- Less efficient than Huffman, but allows one to code symbol s_i with length l_i directly from p_i.

Shannon-Fano Coding Principle

- "A more frequent message has to be encoded by a shorter encoding vector (word) and a less frequent message has to be encoded by a longer encoding vector (word)".
- Uses Top-Down Approach.

Shannon-Fano Encoding: Algorithm

- The letters (messages) of (over) the input alphabet must be arranged in order from most probable to least probable.
- Then the initial set of messages must be divided into two subsets whose total probabilities are as close as possible to being equal. All symbols then have the first digits of their codes assigned; symbols in the first set receive "0" and symbols in the second set receive "1".
- The same process is repeated on those subsets, to determine successive digits of their codes, as long as any sets with more than one member remain.
- When a subset has been reduced to one symbol, this means the symbol's code is complete.

Shannon-Fano Encoding: Example

- Consider a DMS with seven possible symbols s_i, i=1,2,……,7 and the corresponding probabilities p_1=0.35, p_2=0.25, p_3=0.15, p_4=0.08, p_5=0.06, p_6=0.06, and p_7=0.05. Construct the codewords of all the symbols by using Shannon-Fano algorithm. And also determine the following:
  i. Average codeword length
  ii. Entropy of the source
  iii. Coding efficiency
  iv. How much average codeword length exceeds entropy?
  v. Variance of the code.

Variance of the Code

- Variance of the code is given by,

$$\sigma^2 = \sum_{k=0}^{K-1} \left(l_k - \overline{L} \right)^2 p_k$$

Shannon-Fano Encoding: Properties

- It should be taken into account that the **Shannon-Fano code is not unique** because it depends on the partitioning of the input set of messages, which, in turn, is not unique.
- If the successive equiprobable partitioning is not possible at all, the Shannon-Fano code may not be an optimum code, that is, a code that leads to the lowest possible average length of the encoding vector for a given D.