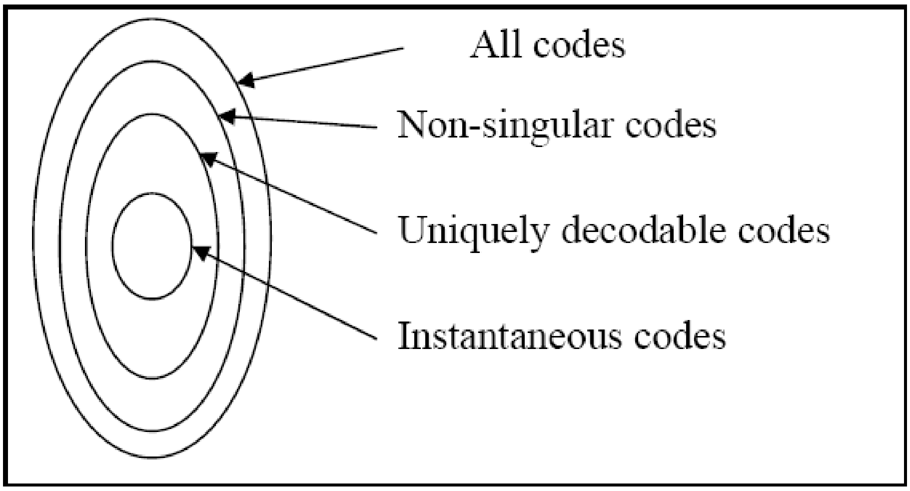
<u>4.2-Coding and Data Compression</u>

Information theory answers two fundamental questions in communication theory:

- what is the ultimate data compression (answer is the *entropy H*), and
- what is the ultimate transmission rate of communication (answer is the *channel capacity C*) .

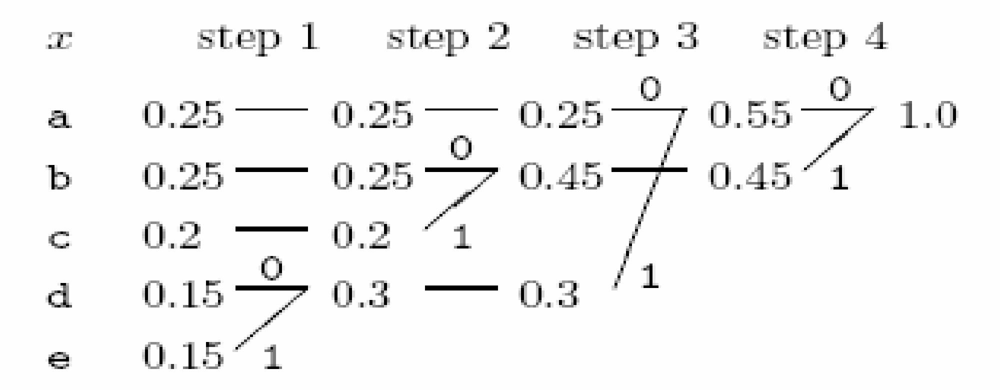
Classes of Codes



Text= aaaaabbbbbccccdddeee

(Huffman coding)

- X ={ a, b, c, d, e }
- P(X) ={ 0.25, 0.25, 0.2, 0.15, 0.15 }.



- The final codes are {00, 10, 11, 010, 011}

a_i	p_i	$h(p_i)$	l_i	$c(a_i)$
a	0.25	2.0	2	00
b	0.25	2.0	2	10
c	0.2	2.3	2	11
d	0.15	2.7	3	010
e	0.15	2.7	3	011

Optimality of Huffman coding

For any distribution, there exists an optimal instantaneous code (with minimum expected length) that satisfies the following properties:

1. a most frequent symbol has length smaller than a less frequent symbol.
2. The two longest codewords have the same length
3. The two longest codewords differ only in the last bit and correspond to the two least likely symbols.

References

Thomas M. Cover and Joy A. Thomas, "Elements of Information Theory", 2nd Ed, 1991