6.

IMAGE CODING

(Source Coding)

- Lossless Coding
  - Entropy coding
  - Huffman code

- Lossy Coding
  - JPEG
  - Bit rate depends on quality

7.

Entropy Coding (Lossless)

Let X_n be a i.i.d sequence of Random Variables taking on values in $\{0, \ldots, M-1\}$

$$P\{X_n = i\} = p_i \qquad 0 \le i \le M-1$$

Entropy

$$H(X_n) \triangleq -\sum_{i=0}^{M-1} p_i \log_2 p_i$$

$$= -E[\log_2 p_i]$$

How do you code RV's into a binary sequence?

Definition: A code is a mapping from the set $\{0, \ldots, M-1\}$ to finite binary sequences.

$M=4$

$$
\begin{aligned}
0 &\rightarrow 01 &= \sigma_0 \\
1 &\rightarrow 10 &= \sigma_1 \\
2 &\rightarrow 0 &= \sigma_2 \\
3 &\rightarrow 10010 &= \sigma_3
\end{aligned}
$$

Definitions

fixed length code -
code in which σ_i have
same length for all i.

Variable length code -
code in which σ_i have
different lengths for
different i's.

9.

Problem

Decode: 0 1 0

CASE 1:

0	1 0
2	1

CASE 2:

0 1	0
0	2

Not Uniquely decodable

Definitions

- A code is <u>Uniquely Decodable</u> if there exist only a single unique decoding.

- A <u>Prefix Code</u> is a specific type of uniquely decodable code in which no code is a prefix of another code.

11.

Theorem: ~~Let~~ C be a uniquely decodable code for the values X_n. Then the ~~average~~ mean code length $\geq H(X_n)$.

$$\bar{n} \triangleq E\left[\,|\sigma_{X_n}|\,\right] = \sum_{i=0}^{M-1} |\sigma_i|\, p_i$$

$$\geq H(X_n)$$

Remember:

X_n are iid

$|\sigma_i|$ = length of code

Question

Can we achieve this bound?

Is the bound tight?

Yes!

Constructive proof

Huffman code

13.

Comments about Huffman Codes

1) Huffman codes are variable length prefix codes.

2) Uniquely decodable

3) Low prob. ⇒ long code
   high prob. ⇒ short code

Algorithm for Constructing a Huffman Code

Step 1: Search the list of probabilities for the two levels with the lowest probabilities. In our example, these are levels 7 and 8 with probabilities 0.04 and 0, respectively.

Step 2: Add the two lowest probabilities together. Form a new list that includes all the probabilities not affected in Step 1, plus the sum generated there. The new list has one less element.

Step 3: Go back to Step 1 and repeat the procedure until the list contains only one element.

Different possible values of $\xi(m,n)$ (source symbols)	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
Huffman code	1	0111	0110	010	001	0001	00001	00000
Probabilities	0.4	0.08	0.08	0.2	0.12	0.08	0.04	0

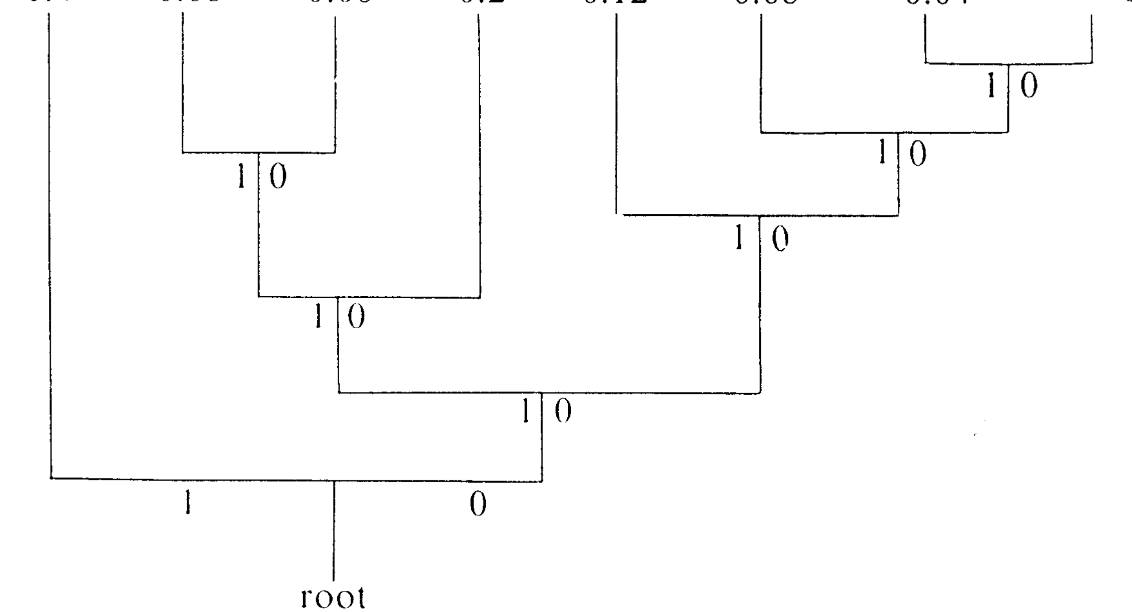

Fig. 26 The tree generated in Fig. 25 is shown here more clearly. This tree is used for determining the code words for the source symbols.

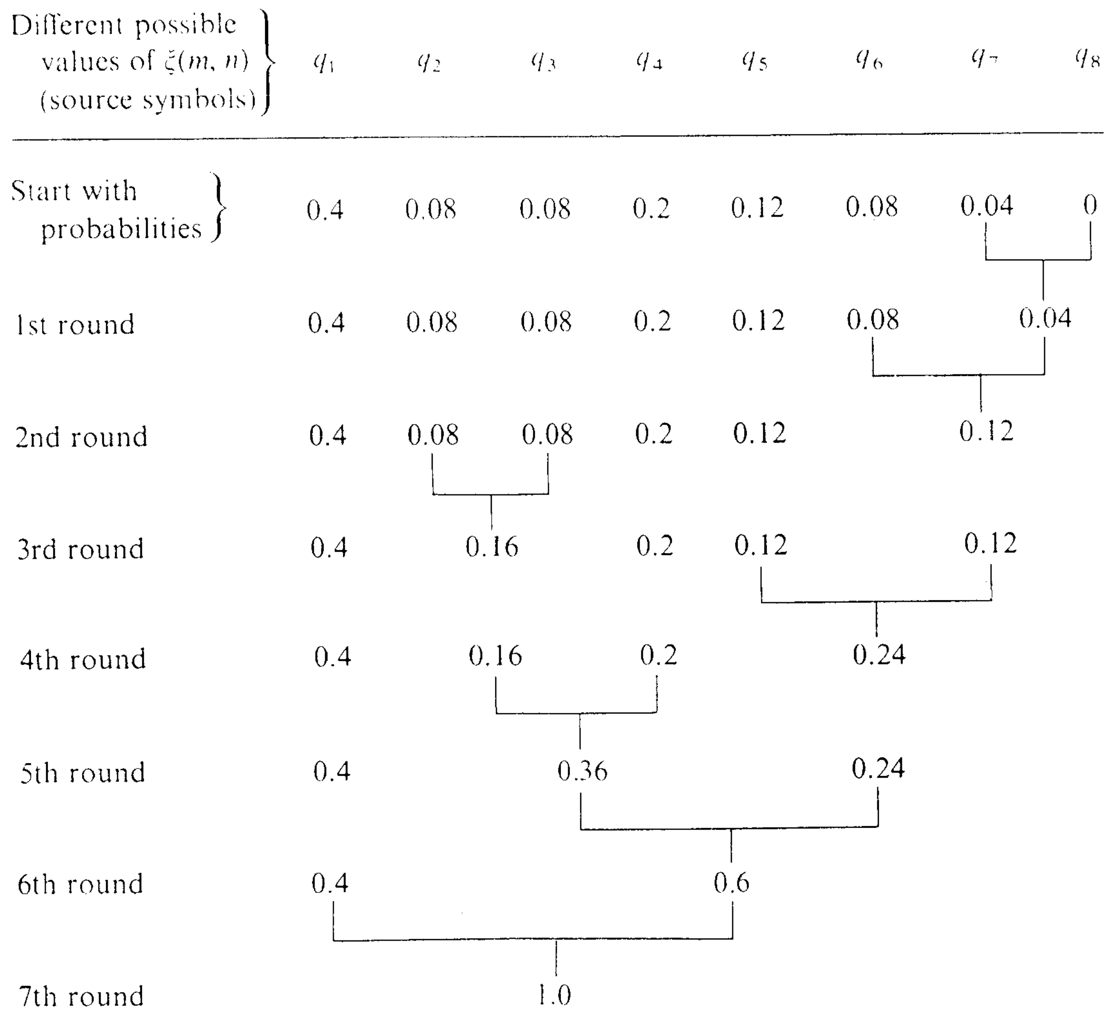

Fig. 25 Building the tree for the Huffman code.

Theorem 1: For a Huffman code, $\bar{n}$ has the property that

$$H(X_n) \le \bar{n} < H(X_n) + 1$$

- Huffman code is within 1 bit of optimal efficiency.
- Can we do better?

Coding in blocks

- Code blocks of the sequence X_n

$$\ldots, \underbrace{X_0, \ldots, X_{m-1}}_{Y_0}, \underbrace{X_m, \ldots, X_{2m-1}}_{Y_1}, \ldots$$

$$Y_n = [X_{nm}, \ldots, X_{nm-1}]$$

$$Y_n \in \{0, \ldots, M^m - 1\}$$

16.

- It is easily shown that

$$H(Y_n) = m\,H(X_n)$$

- By theorem 1):

(*) $\quad H(Y_n) \le \bar{n}_y < H(Y_n) + 1$

$\bar{n}_y$ = mean # of bits per symbol for Y_n.

$\bar{n}_x$ = mean # of bits per symbol for X_n

$(\bar{n}_x = \bar{n}_y / m)$

17.

- Dividing (*) by m results in

$$\frac{1}{m} H(Y_n) \le \frac{\bar{n}_Y}{m} < \frac{H(Y_n)}{m} + \frac{1}{m}$$

$$H(X_n) \le \bar{n}_X < H(X_n) + \frac{1}{m}$$

As $m \to \infty$,

$$\bar{n}_X \approx H(X_n)$$

Comments

1) As the block size m goes to infinity, the bitrate approaches the entropy per input symbol.

$$\lim_{m \to \infty} \bar{n}_x = H(X_n)$$

2) This assume that X_n are i.i.d., but a similar analysis can be applyed to stationary and ergodic sources.