Page No.
Date

Golomb code ($G_m(n)$)

int

- ~~It is used to~~ encode ~~integer~~ value.
- It is used for lossless compression
- Determining Golumb code or codeword
- Assumption:

larger the integer lesser the probability to occure

Algorithm:

step: 1 : unary code of $q = \left\lfloor n/m \right\rfloor$ — integer, m = divisor, ~~devieser~~

step 2 : let $k = \lceil \log_2 m \rceil$

$c = 2^k - m$ and $r = n \bmod m$

$$r' = \begin{cases} r \text{ truncated to } k-1 \text{ bits}, & 0 \le r < c \\ r+c \text{ truncated to } k \text{ bits}, & \text{otherwise} \end{cases}$$

step 3 : concetenate result of set step 1 & step2
// Concetenate (q, r')

* Example: Golomb code

Q: Design Golomb code for 9 with divisor 4,

$n = 9$, $m = 4$, $G_4(9) = ?$

$q = \lfloor 9/4 \rfloor = 2$

Unary code of q = 110 → ①

onary code
2 = 110
3 = 1110
4 = 11110
5 = 111110
n of 1's followed by 0
n of 0's followed by 1

$K = \lceil \log_2 4 \rceil = 2$

$C = 2^K - m$
$= 2^2 - 4$
$= 0$

$r = n \bmod m$
$= 9 \bmod 4$
$= 1$

$r = 1,\ c = 0\ \boxed{r > c}$

$r' = r + c$ truncated to K bits (represented)
$= 0 + 1$ truncated to 2 bits
$= 1$ truncated to 2 bits
$= \underline{\underline{01}}$

Step: 3 concel $\boxed{11001}$

$G_4(9) = 11001$

Entropy: It shows the shortest possible average length of a lossless compression of Data.

Probability $= \dfrac{\text{no. of occurrence}}{\text{total length}}$

Q: Find Entropy:

$P(a_1) = 1/2,\quad P(a_2) = 1/4,\quad P(a_3) = P(a_4) = 1/8$

Entropy =

1 for (1/2) occurrence only

for a_1 $\quad -1\left(\frac{1}{2}\right)\left(\log_2 \frac{1}{2}\right)$

$-1\left(\frac{1}{4}\right)\left(\log_2 \frac{1}{4}\right)$

for a_2 $\quad -2\left(\frac{1}{8}\right)\left(\log_2 \frac{1}{8}\right)$

$= \left(\cancel{-\frac{1}{2})(-1) - \frac{1}{4} - \frac{1}{4}(-2)}\right.$

$= \left(-\frac{1}{2}\right)(-1) - \left(\frac{1}{4}\right)(-2) - \left(\frac{1}{4}\right)(-3)$

$= \frac{1}{2} + \frac{1}{2} + \frac{3}{4} = 1.75$ bits/symbols

* Find entropy if the given sequence is

1 2 1 2 3 3 3 3 1 2 3 3 3 3 1 2 3 3 1 2

Unique characters = 1, 2, 3

$P(1) = 5/20 = 1/4$ $P(2) = 5/20 = 1/4$

$P(3) = 10/20 = 1/2$

$$\text{Entropy} = -2\left(\tfrac{1}{4}\right)\log_2 \tfrac{1}{4} - (1)\tfrac{1}{2}\log_2 \tfrac{1}{2}$$

$= -\frac{1}{2}$ $1 + 0.5 = 1.5$ bits/symbols