2.3 Source Extension (Extension Coding)

2.38. One can usually (not always) achieve better expected length (per source symbol) by encoding blocks of several source symbols.

Definition 2.39. In, an ***n*-th extension** coding, n successive source symbols are grouped into blocks and the encoder operates on the blocks rather than on individual symbols. [4, p. 777]

Example 2.40.

x	$p_X(x)$	Codeword $c(x)$	$\ell(x)$
Y(es)	0.9 (0)	0	1
N(o)	0.1 (1) → 1	1	1

(a) First-order "extension":

$$\mathbb{E}[\ell(X)] = \mathbb{E}[1] = 1 \text{ bit/source symbol}$$

0 1 1 0 0 0 1 0 0 1 1 1

YNNYYYNYYNNN...

Encode two source symbols at a time

(b) Second-order Extension: For DMS, $X_1 \perp\!\!\!\perp X_2$ (independent) — memoryless

x_1x_2 (X_1X_2)	$p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$	$c(x_1,x_2)$	$\ell(x_1,x_2)$
YY	$0.9\times0.9 = 0.81$ (0) → 1	0	1
YN	$0.9\times0.1 = 0.09$ (0) → 0.19 (1)	10	2
NY	$0.1\times0.9 = 0.09$ (0) → 0.1 (1)	110	3
NN	$0.1\times0.1 = 0.01$ (1)	111	3

Let L_n be the expected codeword length per (one) source symbol when Huffman coding is used with n-th extension

$$L_n = \frac{\mathbb{E}[\ell(\underline{X})]}{n}$$

$$\mathbb{E}[\ell(X_1,X_2)] = 1(0.81)+2(0.09)+3(0.09)+3(0.01) = 1.29 \text{ bits per two source symbols}$$

$\underline{X}$ $\quad L_2 = \frac{1.29}{2} = 0.645$ bits per source symbol.

(c) Third-order Extension:

$x_1x_2x_3$	$p_{X_1,X_2,X_3}(x_1,x_2,x_3)$	$c(x_1,x_2,x_3)$	$\ell(x_1,x_2,x_3)$
YYY	$0.9\times0.9\times0.9 = 0.729$		
YYN	$0.9\times0.9\times0.1 = 0.081$		
YNY	$0.9\times0.1\times0.9 = 0.081$		
$\vdots$			
NNN			

$$\mathbb{E}[\ell(X_1,X_2,X_3)] = 1.5980 \text{ bits per 3 source symbols}$$

$\underline{X}$ $\quad L_3 = \frac{1}{3}\times 1.5980 = 0.5327$ bits per source symbol