Digital Communication Systems

EES 452

Asst. Prof. Dr. Prapun Suksompong
prapun@siit.tu.ac.th

2. Source Coding

2.3 Source Extension:

Introduction to Source Extension, Memoryless Source and Independent Symbols

Digital Communication Systems
EES 452

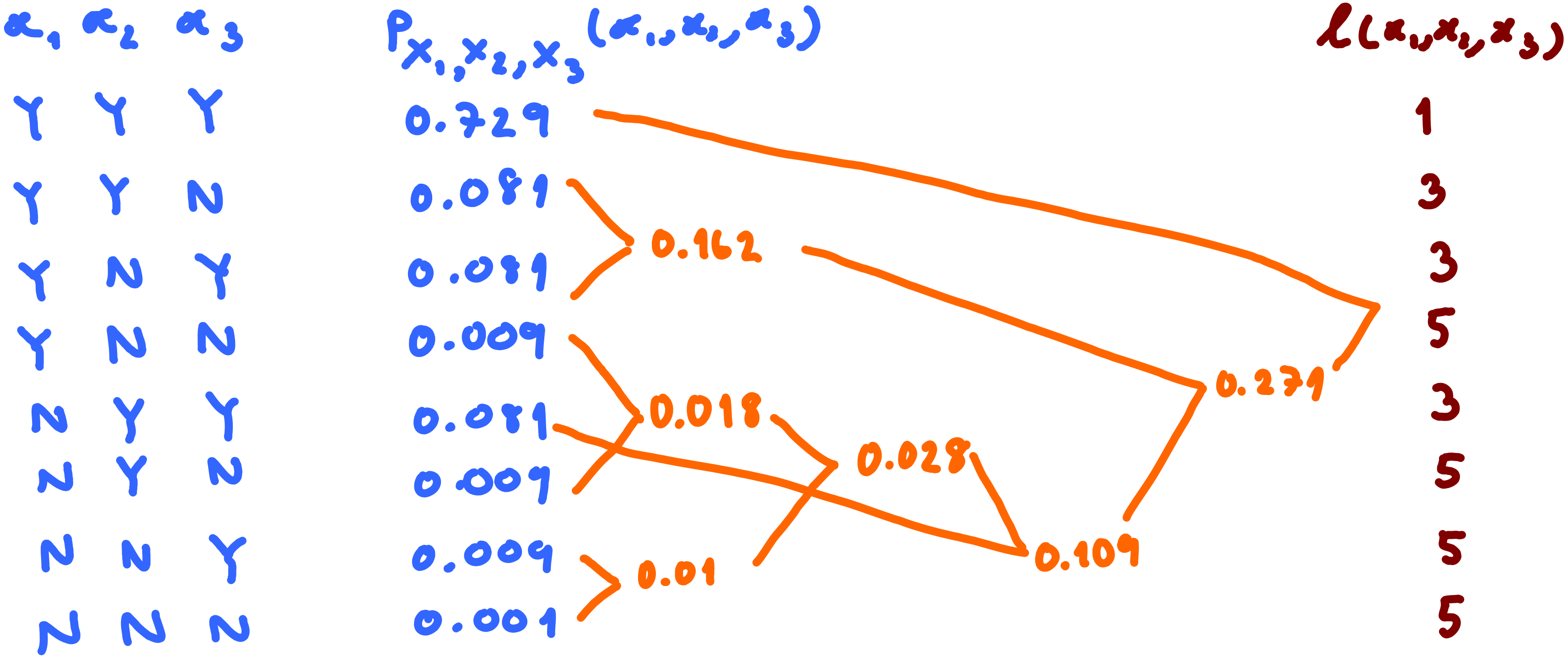
Asst. Prof. Dr. Prapun Suksompong
prapun@siit.tu.ac.th

2. Source Coding

2.3 Source Extension:
Calculating the probabilities for Blocks of Source Symbols, Second- and Third-Order Extensions, L_n

[Ex.2.40]

Huffman Coding: Source Extension



x_1	x_2	x_3	$p_{X_1,X_2,X_3}(x_1,x_2,x_3)$	$\ell(x_1,x_2,x_3)$
Y	Y	Y	0.729	1
Y	Y	N	0.081	3
Y	N	Y	0.081	3
Y	N	N	0.009	5
N	Y	Y	0.081	3
N	Y	N	0.009	5
N	N	Y	0.009	5
N	N	N	0.001	5

Kronecker Product

- An operation on two matrices of arbitrary size
- Named after German mathematician Leopold Kronecker.
- If **A** is an *m*-by-*n* matrix and **B** is a *p*-by-*q* matrix, then the **Kronecker product** **A⊗B** is the *mp*-by-*nq* matrix

Use `kron(A,B)` in MATLAB.

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

- Example

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 1\cdot 0 & 1\cdot 5 & 2\cdot 0 & 2\cdot 5 \\ 1\cdot 6 & 1\cdot 7 & 2\cdot 6 & 2\cdot 7 \\ 3\cdot 0 & 3\cdot 5 & 4\cdot 0 & 4\cdot 5 \\ 3\cdot 6 & 3\cdot 7 & 4\cdot 6 & 4\cdot 7 \end{bmatrix} = \begin{bmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{bmatrix}.$$

Kronecker Product

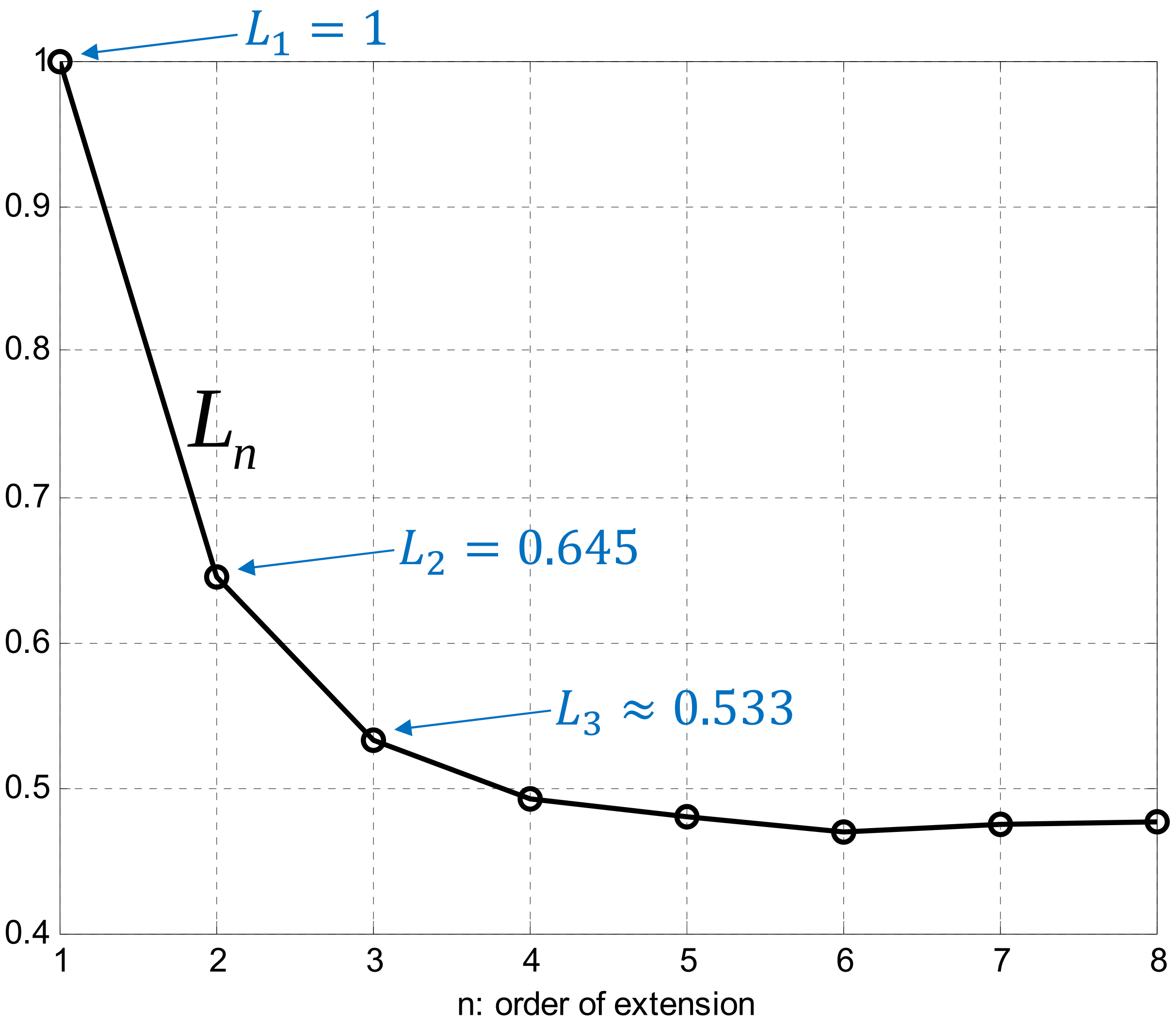
```
>> p = [0.9 0.1]
p =
    0.9000    0.1000
>> p2 = kron(p,p)
p2 =
    0.8100    0.0900    0.0900    0.0100
>> p3 = kron(p2,p)
p3 =
  Columns 1 through 7
    0.7290    0.0810    0.0810    0.0090    0.0810    0.0090    0.0090
  Column 8
    0.0010
```

[Ex.2.40]

Huffman Coding: Source Extension

$$X_k \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$$

$$p = 0.1$$

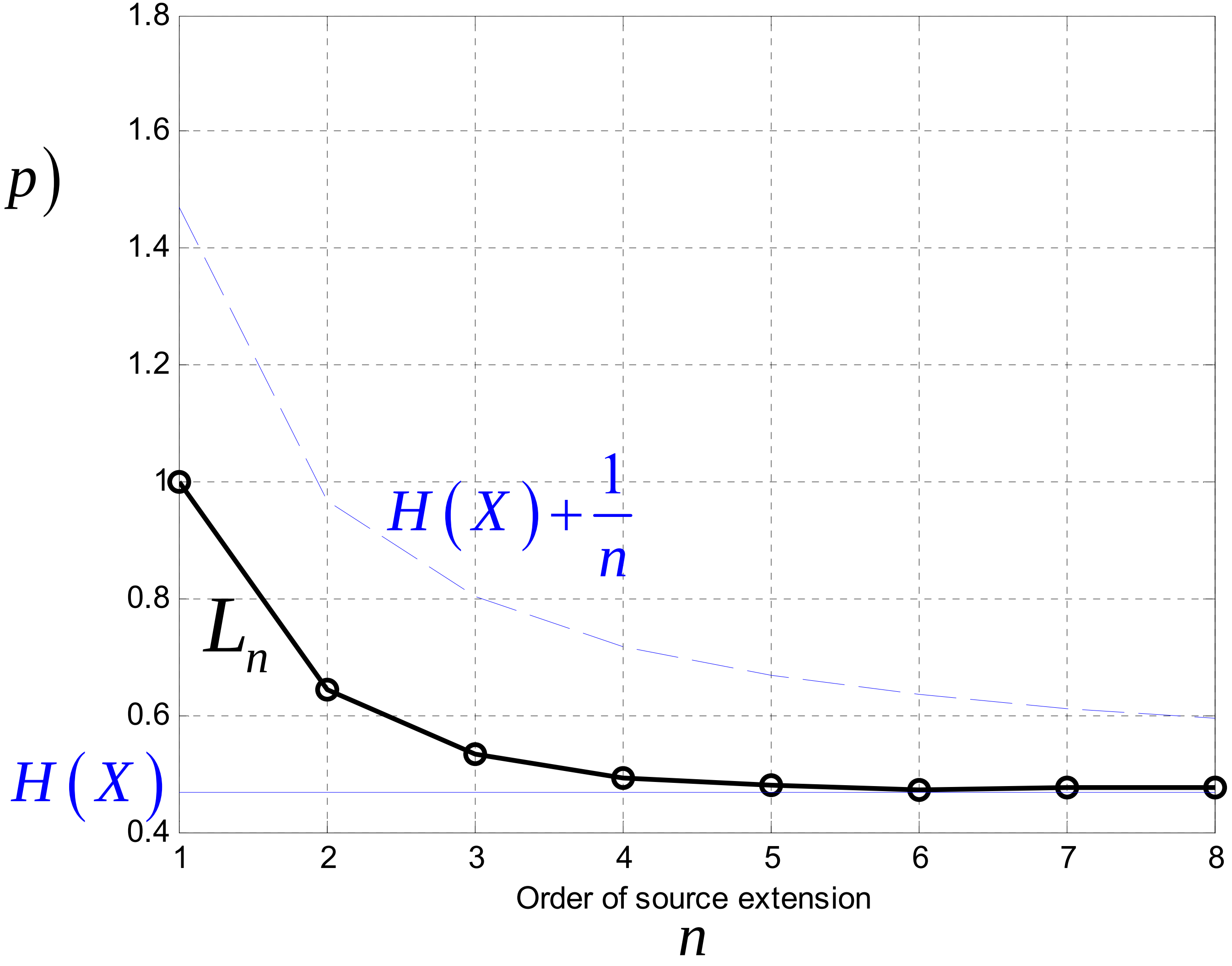


[Ex.2.40]

Huffman Coding: Source Extension

$$X_k \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$$

$$p = 0.1$$

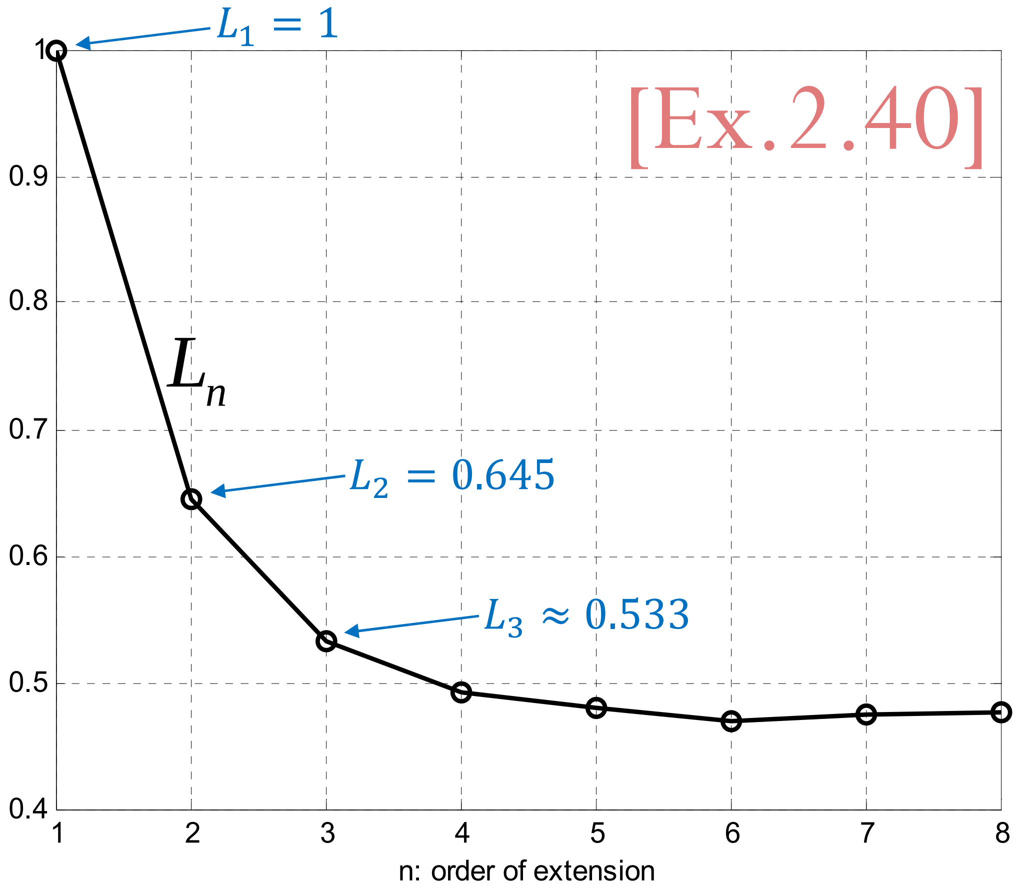


Summary: Source Extension

Source Coding w/o Extension	**Source Coding with Extension**
The encoder operates on individual source symbols.	The encoder operates on blocks of consecutive source symbols. [Defn 2.39] ***n*-th extension coding**: 1 block = n successive source symbols
Source Alphabet $c: \mathcal{A} \to \{0,1\}^*$	**Cartesian Product** $c: (\mathcal{A})^n \to \{0,1\}^*$
Discussed in Sections 2.1-2.2.	Discussed in Section 2.3.
Assuming that the pmf $p(x)$ of the source symbols is known, the optimal encoding technique was discovered by Huffman.	The probabilities for the blocks $p(\underline{\mathbf{x}})$ are calculated from the memoryless property of the source. Huffman's technique can then be applied to the blocks.

Summary: Source Extension

- $\boldsymbol{L_n}$ = expected (average) codeword length <u>per source symbol</u> when Huffman coding is used with n-th extension
    - $\lim_{n\to\infty} L_n = H(X)$

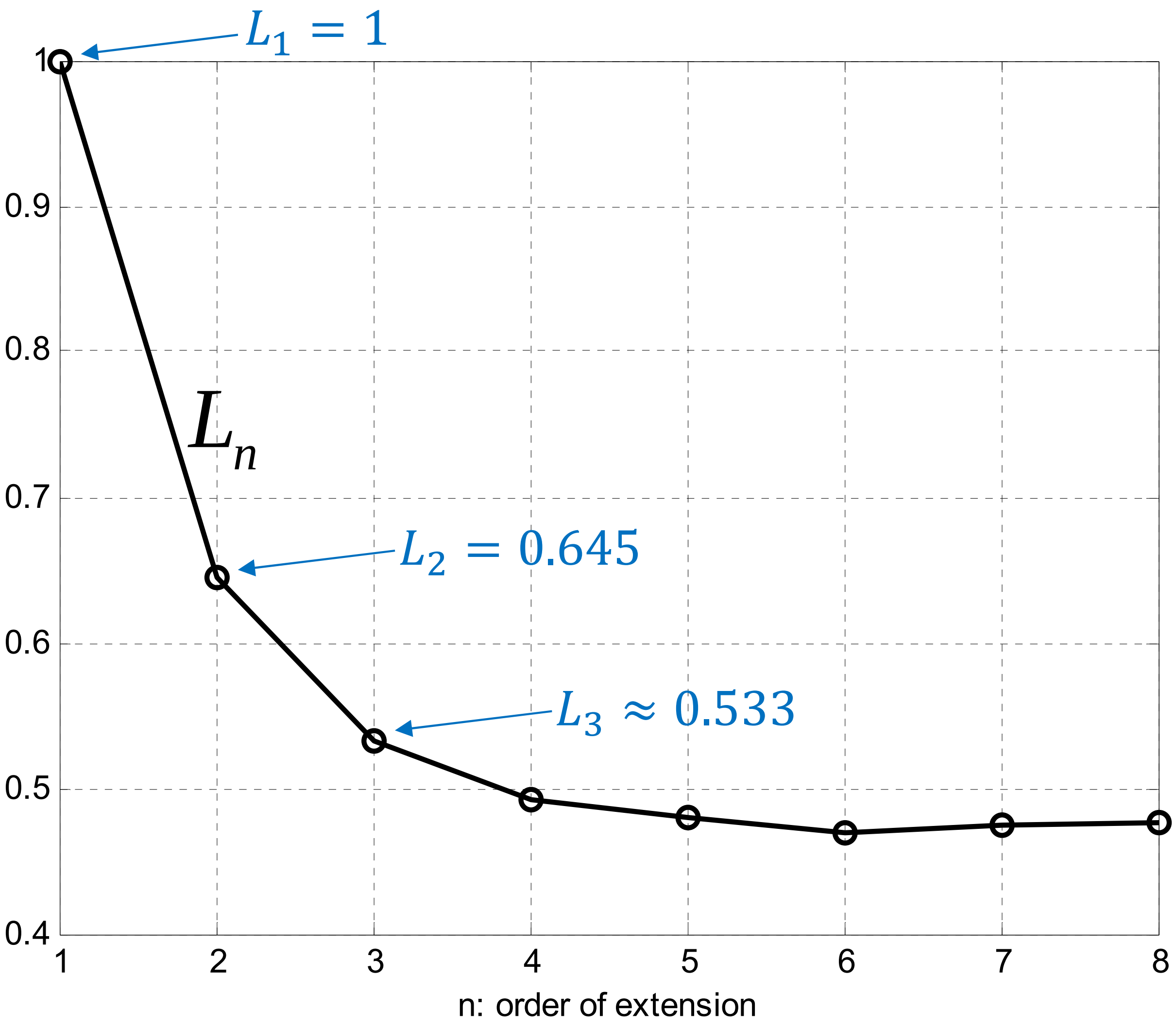


[Ex.2.40]

Huffman Coding: Source Extension

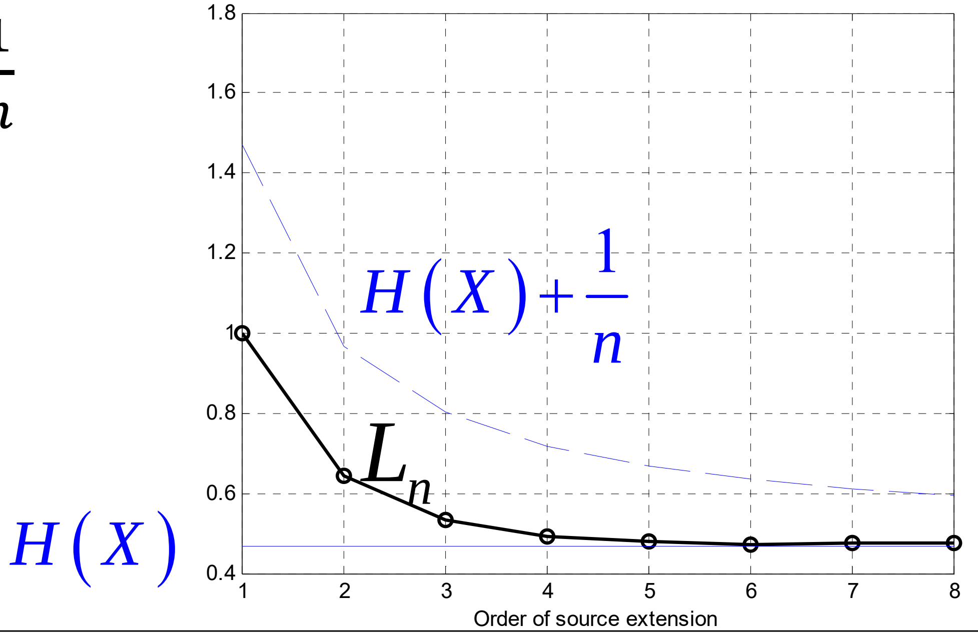
$$X_k \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$$

$$p = 0.1$$



Summary: Source Extension

- The encoder operates on the blocks rather than on individual symbols.
- [Defn 2.39] ***n*-th extension coding**:
  1 block = n successive source symbols
- $\boldsymbol{L_n}$ = expected (average) codeword length per source symbol when Huffman coding is used with n-th extension
  - $H(X) \le L_n < H(X) + \frac{1}{n}$
  - $\lim_{n\to\infty} L_n = H(X)$



Digital Communication Systems

EES 452

Asst. Prof. Dr. Prapun Suksompong
prapun@siit.tu.ac.th

2. Source Coding

Random Vectors

Multiple Random Variables

- Consider a collection of n random variables

$$X_1, X_2, \dots, X_n$$

- Compact notations: **Random Vectors**
  - $\boldsymbol{X}_1^n$
  - $\vec{\boldsymbol{X}} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_n \end{pmatrix}$
  - $\underline{\boldsymbol{X}} = (X_1, X_2, \dots, X_i, \dots X_n)$

Vector Notation

- $\vec{\mathbf{v}}$: column vector $\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix}$
- $\underline{\mathbf{r}}$: row vector $(r_1, r_2, \dots, r_i, \dots r_n)$

$\vec{0}, \underline{0}$: the zero vector
(the all-zero vector)

$\vec{1}, \underline{1}$: the one vector
(the all-one vector)

- Subscripts represent element indices inside individual vectors.
  - v_i and r_i refer to the i^{th} elements inside the vectors $\vec{\mathbf{v}}$ and $\underline{\mathbf{r}}$, respectively.
- When we have a list of vectors, we use superscripts in parentheses as indices of vectors.
  - $\vec{\mathbf{v}}^{(1)}, \vec{\mathbf{v}}^{(2)}, \dots, \vec{\mathbf{v}}^{(M)}$ is a list of M column vectors
  - $\underline{\mathbf{r}}^{(1)}, \underline{\mathbf{r}}^{(2)}, \dots, \underline{\mathbf{r}}^{(M)}$ is a list of M row vectors
  - $\vec{\mathbf{v}}^{(i)}$ and $\underline{\mathbf{r}}^{(i)}$ refer to the i^{th} vectors in the corresponding lists.

Harpoon

- a long, heavy spear attached to a rope, used for killing large fish or whales

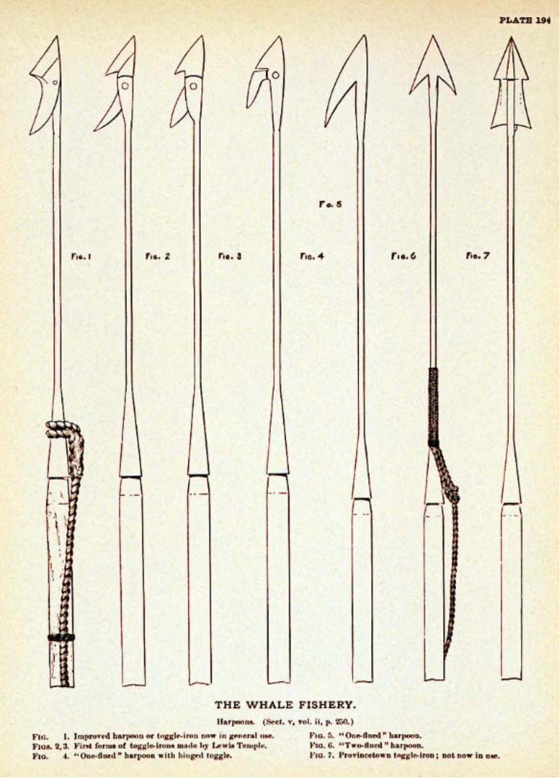


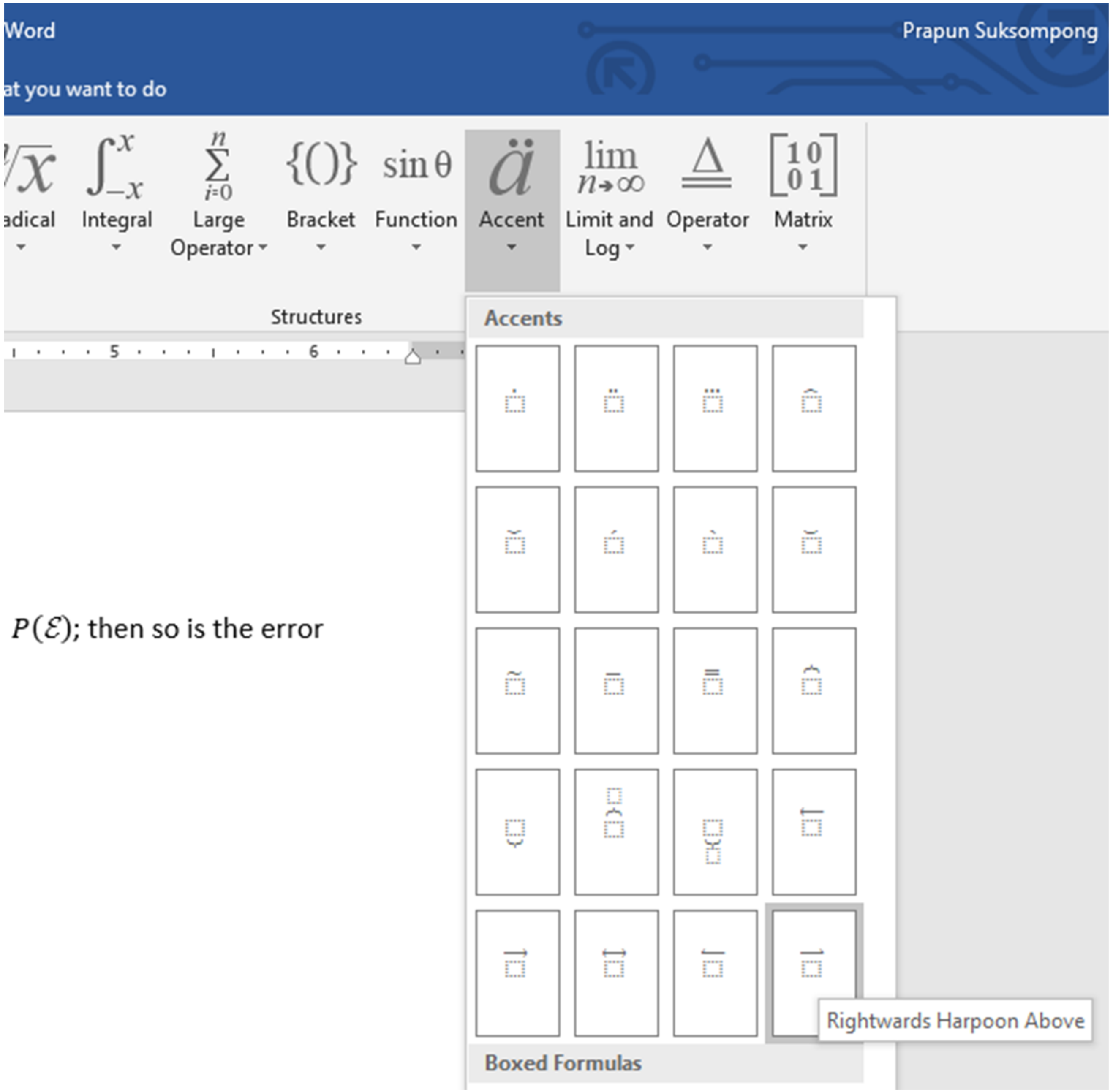
PLATE 194

THE WHALE FISHERY.

Harpoons. (Sect. v, vol. ii, p. 250.)

Fig. 1. Improved harpoon or toggle-iron now in general use.
Figs. 2, 3. First forms of toggle-irons made by Lewis Temple.
Fig. 4. "One-flued" harpoon with hinged toggle.
Fig. 5. "One-flued" harpoon.
Fig. 6. "Two-flued" harpoon.
Fig. 7. Provincetown toggle-iron; not now in use.

Harpoon



$P(\mathcal{E})$; then so is the error

Joint pmf

$$\underbrace{p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)}_{p_{\underline{\boldsymbol{X}}}\left(\underline{\boldsymbol{x}}\right)} = \underbrace{P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]}_{P\left[\underline{\boldsymbol{X}} = \underline{\boldsymbol{x}}\right]}$$

$$p_{\underline{\boldsymbol{X}}}\left(\underline{\boldsymbol{x}}\right) = P\left[\underline{\boldsymbol{X}} = \underline{\boldsymbol{x}}\right]$$

Independence

n random variables $X_1, X_2, \dots, X_n$

are **independent** if for any $x_1, x_2, \dots, x_n$,

$$\underbrace{p_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)}_{p_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{x}})} = \underbrace{p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n)}_{\prod_{\boldsymbol{i}} p_{X_i}(x_i)}$$

$$p_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{x}}) = \prod_{\boldsymbol{i}} p_{X_i}(x_i)$$

Independence

n random variables $X_1, X_2, \dots, X_n$

are **independent** if for any $x_1, x_2, \dots, x_n$,

$$\underbrace{p_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)}_{p_{\boldsymbol{X}_1^n}(\boldsymbol{x}_1^n)} = \underbrace{p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n)}_{\prod_{i=1}^{n} p_{X_i}(x_i)}$$

$$p_{\boldsymbol{X}_1^n}(\boldsymbol{x}_1^n) = \prod_{i=1}^{n} p_{X_i}(x_i)$$

Extension Coding

List Notation	Vector Notation
$p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)$	$p_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{x}})$ or $p_{\boldsymbol{X}_1^n}(\boldsymbol{x}_1^n)$
$c(x_1,x_2,\ldots,x_n)$	$c(\underline{\boldsymbol{x}})$ or $c(\boldsymbol{x}_1^n)$
$\ell(x_1,x_2,\ldots,x_n)$	$\ell(\underline{\boldsymbol{x}})$ or $\ell(\boldsymbol{x}_1^n)$
$\mathbb{E}[\ell(X_1,X_2,\ldots,X_n)]$	$\mathbb{E}[\ell(\underline{\boldsymbol{X}})]$ or $\mathbb{E}[\ell(\boldsymbol{X}_1^n)]$
$L_n = \frac{\mathbb{E}[\ell(X_1,X_2,\ldots,X_n)]}{n}$	$L_n = \frac{1}{n}\mathbb{E}[\ell(\underline{\boldsymbol{X}})]$ or $\frac{1}{n}\mathbb{E}[\ell(\boldsymbol{X}_1^n)]$