A Recap of Last Few Sessions:

i) WLLN: The average M_N of N realizations of a RV X converges, as $N \to \infty$, to $E[X]$

ii) The WLLN implies The Typical set.

As $N \to \infty$, The random sequence $\{X_1, X_2, \ldots, X_N\}$ has to be a member of The Typical set:

→ All The ~~set~~ members have an identical probability. For Bernoulli(p)

1 RV X, where $P(X=1) = p$ and $Prob(X=0) = 1-p$, $P_i = p^{Np}(1-p)^{N(1-p)}$

$$= 2^{-NH_b(p)}$$

where $H_b(p)$ is the Binary Entropy function: $H_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$

→ The size of the typical set, i.e., the cardinality of the typical set, is $\frac{1}{P_i} = 2^{N H_b(p)}$.

→ If we have a set with $2^2 = 4$ members all of which have the same probabitity, we need 2 bits to
2 represent any of these four members. If 2^3 members, 3 bits are needed, and so on, i.e., we can use fixed length codes. Therefore, we can use a fixed length code, whose codewords are of length $N H_b(p)$ bits, to represent

any member of the typical set.

Thus, we have reduced the number of bits from N to $N H_b(p)$ since $H_b(p) \in [0, 1]$.

3

iii) Encoding schemes for DMS:

→ Fixed length codes (FLC)

→ Variable length codes (VLC)

We want to use the VLC when the PMF $\{p_l\}$ corresponding to the source alphabet $\{x_l\}$, $1 \leq l \leq L$, is not uniform.

An example is the Morse Code, which assigns short codes to frequently-occuring letters of English alphabet.

Suppose $X \sim$ Bernouilli (p)

Each member of The typical set,

has to have $\underline{Np}$ ones, and

$\underline{N \cdot (1-p)}$ zeros, Therefore its prob.

$$P_i = p^{Np} (1-p)^{N(1-p)}$$

4

$$= 2^{\log_2 \left(p^{Np} (1-p)^{N(1-p)} \right)}$$

$$= 2^{N \underbrace{\left(p \log_2 p + (1-p) \log_2 (1-p) \right)}_{-H_b(p)}}$$

$$= 2^{-N H_b(p)}$$

Due to the WLLN,

$$\text{Prob}(\text{Typical set}) = 1$$

Denoting the cardinality of the the typical set as K,

$$\Rightarrow \quad K \, P_i = 1 \qquad (5)$$

$$\Rightarrow \quad K = \frac{1}{P_i}$$

$$= 2^{N H_b(P)}$$

We have reduced the required number of bits from N to $NH_b(P)$

We would like to keep the VLC

a prefix-free code.

Today: An important statement of

the Information Theory, called the

Kraft's Inequality: 6

→ Connects the two of the earlier

topics, the Entropy function, and

the coding scheme (FLC / VLC)

$$\boxed{\text{If the DMS symbols } \{x_\ell\} \text{ are encoded using } \{n_\ell\} \text{ bits using the prefix-free code, then } \sum_{\ell=1}^{L} 2^{-n_\ell} \leq 1}$$

Examples :

The DMS size $L = 8$,

and we have used the FLC.

Therefore $n_l = 3$ for $l = 1, \dots, 8$.

$$\sum_{l=1}^{L=8} 2^{-n_l} = \sum_{l=1}^{8} 2^{-3} = 1.$$

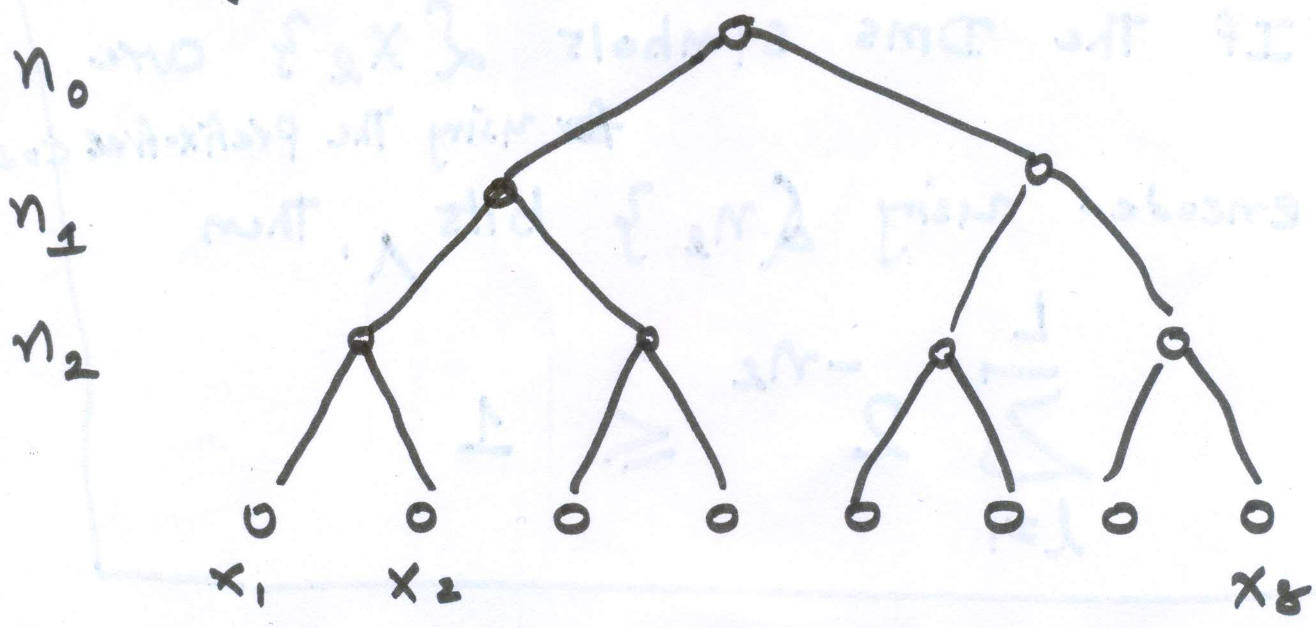
An observation: we use the FLC

if all values of P_l are the same, i.e.,

$P_l = 2^{-n_l}$.

7

A graphical viewpoint:



Another Example:

Let us consider a prefix-free VLC for a source alphabet $\{x_1, x_2, x_3, x_4\}$

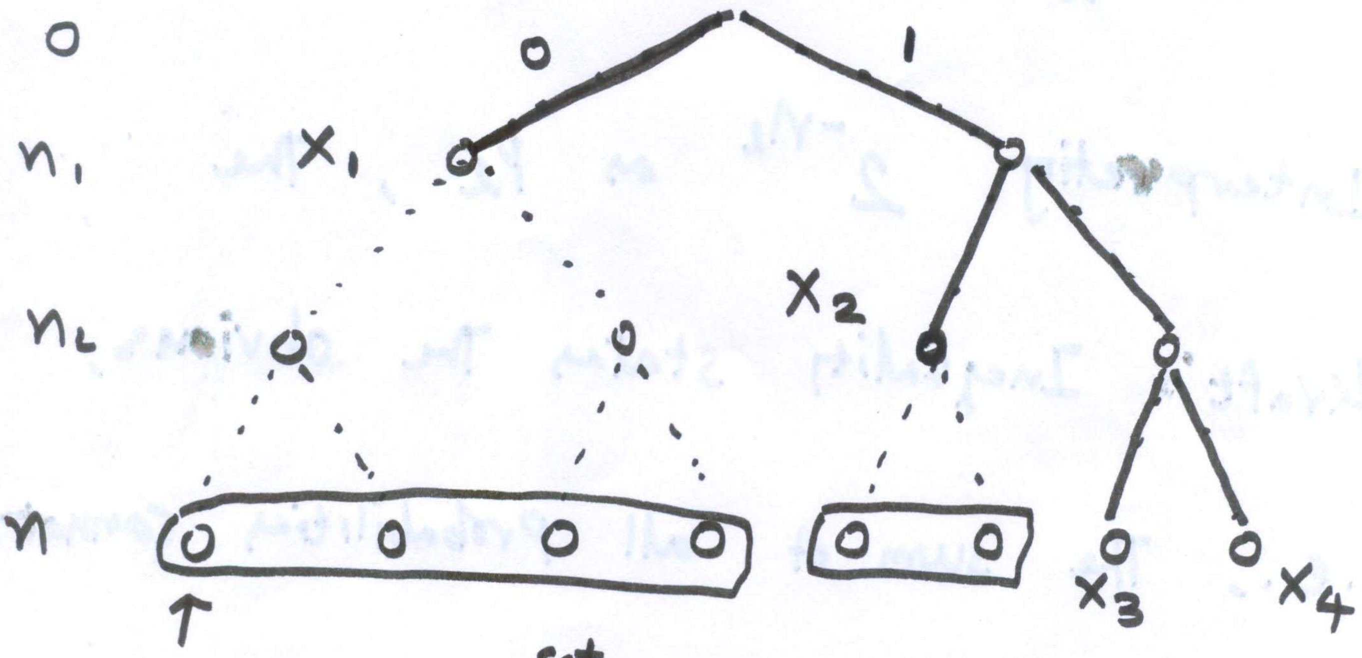
$n_1 = 1$ bit and $x_1 : 0$

$n_2 = 2$ bits. $x_2 : 10$

$n_3 = n_4 = 3$ bits $x_3 : 110$

$x_4 : 111$

8



The size of the set leaf nodes that get ruled out by assignment of x_1 as $A_1 = 4 = 2^{n-n_1} = 2^{3-1}$

$A_2 = 2 = 2^{3-2} = 2^{n-n_2}$

$A_3 = A_4 = 1$

$2^{3-3} = 2^{n-n_3}$

For the prefix-free codes, the following has to hold:

$$\sum_{l=1}^{L} A_l \leq 2^n$$

$$\Rightarrow \sum_{l=1}^{L} 2^{n-n_l} \leq 2^n$$

$$\Rightarrow \sum_{l=1}^{L} 2^{-n_l} \leq 1$$

9

Interpreting 2^{-n_l} as P_l, the Kraft's Inequality states the obvious, i.e., the sum of all probabilities cannot exceed 1.

Now, if $2^{-n_l} = P_l$, then

$$n_l = -\log_2(P_l) = \log_2\left(\frac{1}{P_l}\right)$$

Define $I_l = \log_2 (1/p_l)$ bits.

as the Information generated

by the occurrence of x_l

what is the ~~average~~ Expected value of the information

generated by the random X ?

10

$$\sum I_l \cdot p_l$$

$$= \sum p_l (-\log_2 (p_l))$$

$$= H(X)$$

The Entropy is defined

for Bernouilli (p) RV X

$$P(X=1) = p = P_1$$

$$P(X=0) = 1-p = P_0$$

$$H_b(p) = H(X) = -(\log_2 p)\,p - (\log_2(1-p))(1-p)$$

$$= -P_1 \log_2 P_1 - P_0 \log_2 P_0$$

This definition generalizes to the following when RV X takes one of L symbols

$$H(X) = \sum_{l=1}^{L} P_l \cdot (-\log_2(P_l))$$