17 March, 2020 14:13

X r.v.

$C : X \to \{0,1\}^*$

$x \in X \qquad C(x)$

$\ell(x) = |C(x)|$ ... length of encoding

Q: $\mathbb{E}[\ell(x)] \approx H(X)$ ... for good encoding

$y \in \{C(x),\ x \in X\}$ ... code $= C$

↳ code word

C^* closure of code C

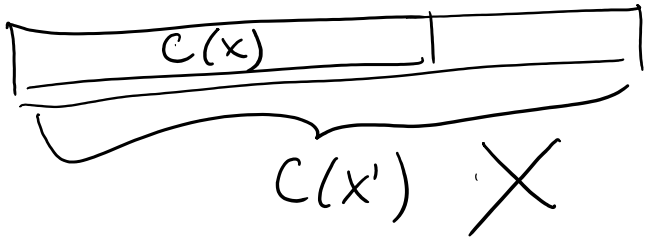
$= \{y_1 y_2 \cdots y_k,\ y_1, \cdots y_k \in C\}$

C unambiguous ... if $\forall x_1, \cdots x_k \in X$, $x'_1, \cdots x'_{k'}$

$$C(x_1)C(x_2)\cdots C(x_k) \neq C(x'_1)C(x'_2)\cdots C(x'_{k'})$$

C prefix-free if $\forall x, x' \in X$

$C(x)$ is not a prefix of $C(x')$



- C prefix free $\Rightarrow$ C unambiguous

$\not\Rightarrow$ (Exc)

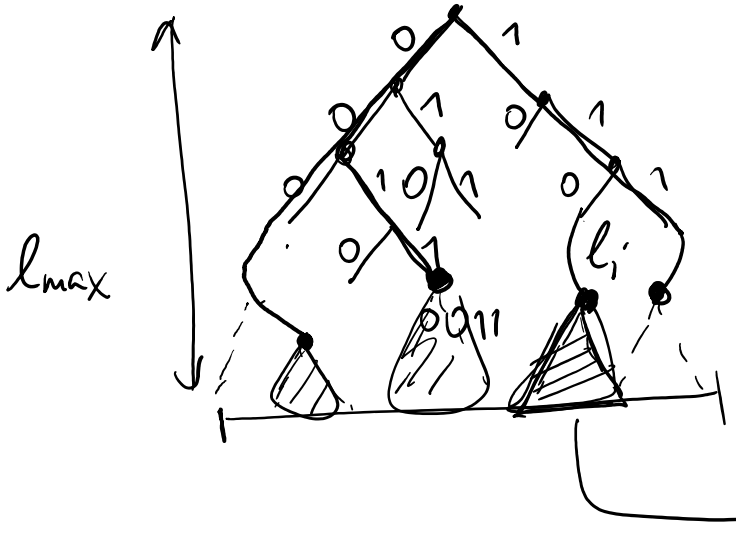
C suffix-free $\Rightarrow$ ...

0 1
0

Thm: (Kraft Inequality)
Let C be prefix-free code with code-lengths $l_1, l_2, \ldots$

$$\sum_i 2^{-l_i} \leq 1$$

Pf:



C is finite

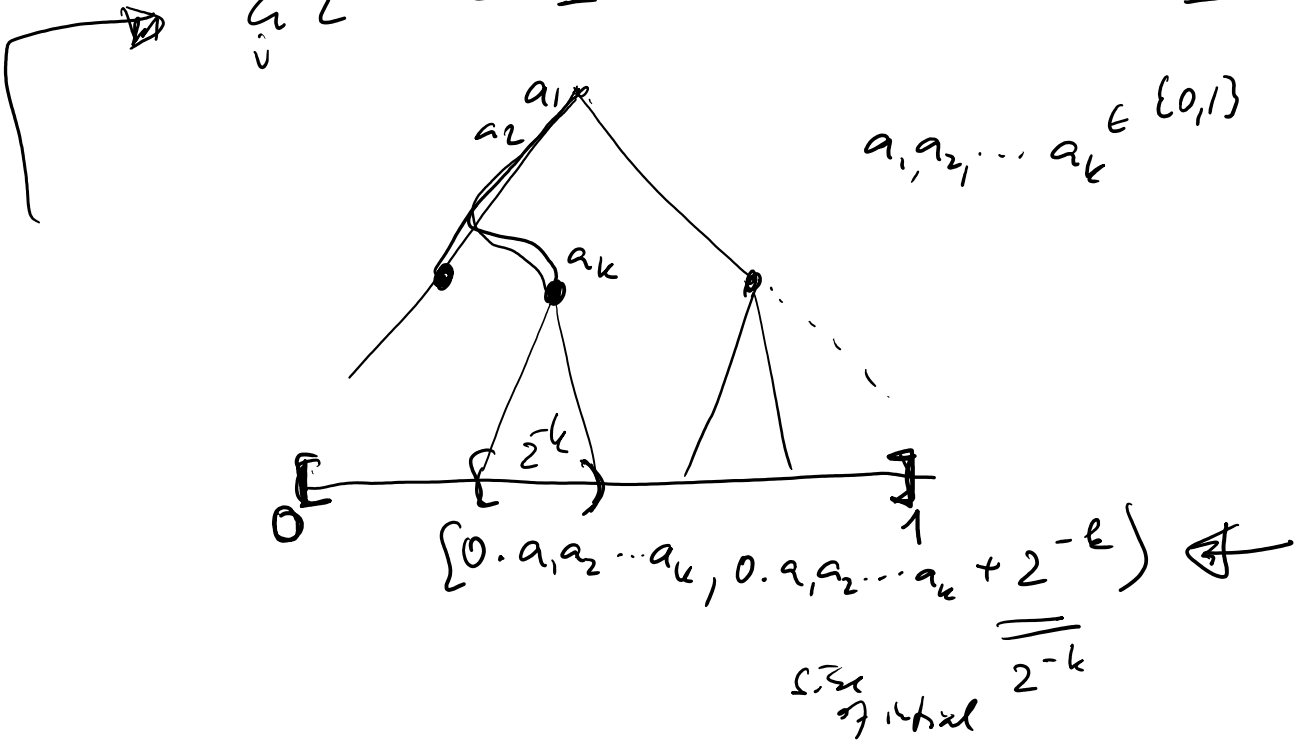
$\rightarrow 2^{l_{max} - l_i}$

$2^{l_{max}}$ clear

$$\sum_i 2^{l_{max} - l_i} \leq 2^{l_{max}}$$

$$2^{l_{max}} \sum_i 2^{} \leq 2^{l_{max}}$$

$$\sum_i 2^{-l_i} \leq 1$$ ✓3

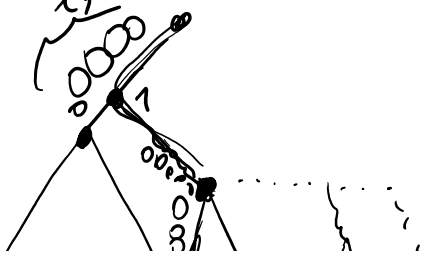


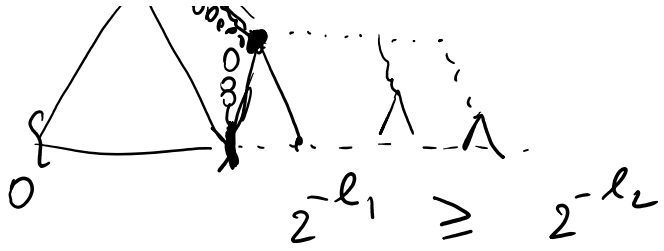
$a_1, a_2, \ldots a_k \in \{0,1\}$

Claim If $l_1, l_2, \ldots$ satisfy $\sum 2^{-l_i} \leq 1$

$\Rightarrow \exists$ code C prefix-free with lengths $l_1, l_2, \ldots$

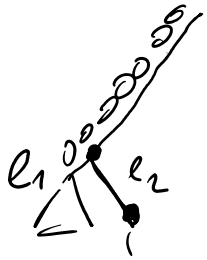
Pf: $l_1 \leq l_2 \leq \cdots$ sort



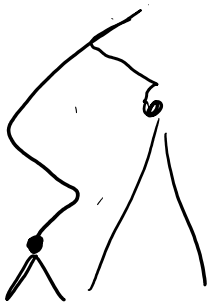


Starting place for interval l_i
is a multiple of l_i

$\simeq$ depth-first search



- sort codewords lexicographically



- optimal <u>prefix-free</u> code C for r.v. X

$C : X \to \{0,1\}^*$ $\qquad L = \mathbb{E}[|C(x)|]$

$$L = \sum_{x \in X} p(x) \cdot l(x)$$

$$L - H(X) \geq 0$$

$$L - H(X) = \sum_{x \in X} p(x) \cdot l(x) - \sum_{x \in X} p(x) \cdot \lg \frac{1}{p(x)}$$

$2^{l(x)}$ ←

$$= \sum_{x \in X} p(x) \lg \frac{2^{\ell(x)} \cdot p(x)}{\cdot} = (*) \geq 0$$

want

$$c = \sum 2^{-\ell(x)} \leq 1 \qquad q(x) = \frac{2^{-\ell(x)}}{c}$$

Kraft

$$\sum_x q(x) = 1$$

$$(*) = \sum_{x \in X} p(x) \lg p(x) \cdot 2^{+\ell(x)} \qquad D(p \| q) \geq 0$$

$$= \sum p(x) \lg \frac{p(x)}{c \cdot q(x)} = \sum p(x) \cdot \lg \frac{p(x)}{q(x)} + \sum p(x) \lg \frac{1}{c} \geq 0$$

$\lg \frac{1}{c} \geq 0$ for $c \leq 1$

$$L - H(x) \geq 0$$

$$\Rightarrow L \geq H(x) \qquad \square$$

Shannon code:

$$\ell(x) = \lceil \lg \frac{1}{p(x)} \rceil$$

Ex: $p(x) = 2^{-10} \rightarrow \ell(x) = 10$

$$\sum 2^{-\ell(x)} \leq 1$$

$$\sum 2^{-\lceil \lg \frac{1}{p(x)} \rceil} = \sum \frac{1}{2^{\lceil \lg \frac{1}{p(x)} \rceil}} \leq \sum \frac{1}{2^{\lg \frac{1}{p(x)}}}$$

$$= \sum \frac{1}{\frac{1}{\ldots}}$$

$\hookrightarrow \frac{1}{p(x)}$

$\Rightarrow \exists$ prefix-free $= \sum p(x) = 1$

code for $\ell(x)$'s.

$$\underbrace{\sum p(x) \lg \frac{1}{p(x)}}_{H(x)} \leq \sum p(x) \cdot \ell(x) \leq \underbrace{\sum p(x) \left[\left(\lg \frac{1}{p(x)} \right) + 1 \right]}_{H(x) \quad + 1}$$

$$H(x) \leq L \leq H(x) + 1$$ ☒

optimal upto one bit!

Ex: 1) $p(x_1) = 0.000001$ $\ell(x_1) = 20$ $\overbrace{1\,0\,0\,0\,0\,0\,0}^{10}$

$p(x_2) = 0.999999$ $\ell(x_2) = 1$ 0

2) $p(x_1) = \frac{1}{3}$

$p(x_2) = \frac{1}{3}$ $\ell(x_1) = \ell(x_2) = \ell(x_3) = 2$

$p(x_3) = \frac{1}{3}$ opt: 0, 10, 11

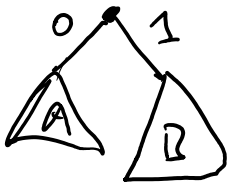
Huffman code: ← provably optimal ☒

$p_1 \geq p_2 \geq \dots \geq p_{n-1} \geq p_n$

$p_{n-1} + p_n$

0 ∧ 1

- $p_{n-1} \geq p_{n-2}$



Fano code: preserves order of $p_1, p_2, \dots p_n$

$$\left\{ \underbrace{p_1 + p_2 + \dots p_k}_{} \;\middle|\; \underbrace{p_{k+1} + p_{k+2} + \dots p_n}_{} \right\}$$

minimize difference

$$\min_k \left| \sum_{i=1}^{k} p_i - \sum_{i=k+1}^{n} p_i \right|$$

proceed recursively

Fact: $L \leq H(X) + 2$

- 1 bit loss might be too large

pick k build a code for $(X_1, X_2 \dots, X_k)$

$(X_1, X_2 \dots, X_k)$

$$H(X_1 \dots X_k) = k\, H(X)$$

$$L_k = H(X_1 \dots X_k) + 1$$

losing $\frac{1}{k}$ bits per r.v.

Thm: (McMillan): For any unambiguous C with code lengths $l_1, l_2, \dots \leq l_{max}$

$$\sum 2^{-l_i} \leq 1 \quad .$$

Pf C^k ... concatenations of k words from C

$$\left(\sum_{x \in C} 2^{-\ell(x)}\right)^k = \sum_{x_1 \in C} \dots \sum_{x_k \in C} 2^{-\ell(x_1) - \ell(x) \dots \ell(x_k)}$$

$$\sum_{\bar{x} \in C^k} 2^{-\ell(\bar{x})} = \sum_{m=1}^{k \cdot l_{max}} w(m) \cdot 2^{-m} = (*)$$

$w(m)$ ... # of words of length m in C^k

$$w(m) \leq 2^m$$

$$(*) \leq \sum_{m=1}^{k \cdot l_{max}} 2^m \cdot 2^{-m} = k \cdot l_{max}$$

$k \geq 1$

$$\Rightarrow \left(\sum_{x \in C} 2^{-\ell(x)} \right)^k \leq k \cdot \ell_{max}$$

$$\sum_{x \in C} 2^{-\ell(x)} \leq \left(k \cdot \ell_{max} \right)^{\frac{1}{k}}$$

fixed

$$\longrightarrow 1 \quad (k \to \infty)$$

$$\Rightarrow \sum_{x \in C} 2^{-\ell(x)} \leq 1 \qquad \square$$

Q: C code (not necessarily prefix free)

X r.v.

$\mathbb{E}[|C(x)|]$

HW ?