Lec 11 - Krafts Inequality

Sep 23^{rd} 2017

Recap

- Source Code $\mathcal{C}$
- $c(x)$, $\ell(x)$
- $L(\mathcal{C})$
- examples
- nonsingular
- extension
- uniquely decodable
- prefix code
- KI, existence of a prefix code over alphabet $\mathbb{D}$, $|\mathbb{D}| = D$

$$\iff \sum_{i=1}^{m} D^{-\ell_i} \leq 1$$

$\mathbb{D} = \{0, 1, \dots, D-1\}$

Today.

1. Extended KI
2. $L \geq H_D(X)$
3. $H_D(X) \leq L^* \leq H_D(X) + 1$
4. $L_n^* \to H(\mathcal{X}_0)$
5. Wrong code $D(p\|q)$ comes in.

Theorem (Extended Krafts Inequality)

For any countably infinite set of codewords that form a prefix code, we have the extended K.I.

$$\sum_{i=1}^{\infty} D^{-\ell_i} \leq 1 \quad \text{--- } (1)$$

Conversely given any $\{\ell_i\}$ satisfying (1), we can find a prefix code with lengths that satisfy (1).

Proof.

Let ℓ be a code having a countably ∞ set of (source) codewords of length $\{\ell_i\}_{i=1}^{\infty}$. Suppose

$$\underset{\text{codeword}}{i^{th}} \iff y_1 y_2 \cdots y_{\ell_i}$$

Recall that in a prefix Code, no codeword can be a prefix of any other codeword. We associate with codeword $y_1 y_2 \cdots y_{l_i}$ the real number $0.y_1 \cdots y_{l_i}$

if $D=2$, $\mathbb{D}=\{0,1\}$
$c(x)=101$, then $0.101=\frac{1}{2}+\frac{1}{8}$

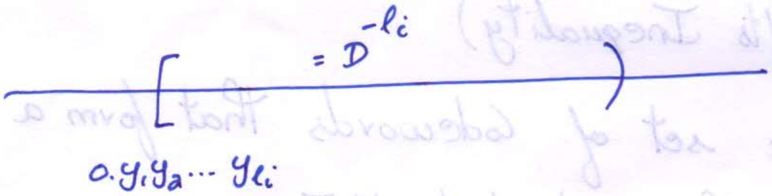
It follows that no other codeword is associated with real number in the interval

$$\left[0.y_1 \cdots y_{l_i},\ 0.y_1 \cdots y_{l_i}+D^{-l_i}\right)$$

Thus all the intervals must be disjoint. The length of this interval is D^{-l_i}.

It follows that

$$\sum_{i=1}^{\infty} D^{-l_i} \leq 1$$



Conversly suppose one is given lengths satisfying,

$$\sum_{i=1}^{\infty} D^{-l_i} \leq 1$$

wolog assume,

$$l_1 \leq l_2 \leq \cdots\cdots$$

Associate with l_1 the interval $[0, D^{-l_1})$ and assign codeword $0.\underbrace{000\cdots0}_{l_1 \text{ zeros}}$

Associate with l_2 the interval $\left[D^{-l_1}, D^{-l_1}+D^{-l_2}\right)$, and the codeword D^{-l_1} etc.

In this way, we see that the prefix condition is satisfied (the intervals are disjoint, hence no codeword is prefix of any other codeword).

x———————x

Optimal Codes-

Assume $|\mathcal{X}_0| = m$

Pose the problem as one of Constraint minimisation.

$$\min \sum_{i=1}^{m} P_i l_i \quad , \text{subject to} \quad \sum_{i=1}^{m} D^{-l_i} \leq 1$$

(we also want l_i to be integer, but ignore this (RELAXATION))

Use "lagrenge-multiplier" to solve the problem

$$\frac{\partial}{\partial l_i} \left(\sum_{i=1}^{m} P_i l_i + \lambda \sum_{i=1}^{m} D^{-l_i} \right) = 0 \quad \text{—— } 1$$

$$\sum_{i=1}^{m} D^{-l_i} = 1$$

Note- if $\sum_{i=1}^{m} D^{-l_i} < 1$, can improve minimisation.

Rewritting ①,

$$\frac{\partial}{\partial l_i}\left(\sum_{i=1}^{m} p_i l_i + \lambda \sum_{i=1}^{m} e^{-l_i a}\right) = 0 \quad \text{, where} \quad a = \log_e D$$

$$\Rightarrow \quad p_i + \lambda(-a)e^{-l_i a} = 0$$

$$\therefore \quad D^{-l_i} = \frac{p_i}{\lambda a}$$

$$\therefore \quad \sum_{i=1}^{m} D^{-l_i} = \sum_{i=1}^{m} \frac{p_i}{\lambda a} = \frac{1}{\lambda a}$$

$$\therefore \quad \lambda a = 1$$

$$\therefore \quad D^{-l_i} = p_i$$

$$\Rightarrow \quad \boxed{l_i = \log_D \frac{1}{p_i}}$$

This suggests that integer asignment,

$$l_i = \left\lceil \log_D \frac{1}{p_i} \right\rceil$$

with this we get,

$$\sum_{i=1}^{m} p_i l_i < \sum_{i=1}^{m} p_i \left\lceil \log_D \frac{1}{p_i} \right\rceil$$

$$= \underbrace{\sum_{i=1}^{m} p_i \log \frac{1}{p_i}}_{H_D(x)} + 1$$

$$= H_D(x) + 1$$

$$\therefore \quad \boxed{L(\mathcal{C}) < H_D(x) + 1}$$

Theorem

The expected length of any instantaneous D-ary code for a RV X satisfies
(prefix)

$$L \geq H_D(x)$$

with equality if $p_i = D^{-l_i}$ (D-adic source)

Proof-

Define $q_i = \dfrac{D^{-l_i}}{\sum_{i=1}^{m} D^{-l_i}}$, Then $\sum_{i=1}^{m} q_i = 1$

Consider

$$\cancel{L - H_D(x)} \quad L - H_D(x) = \sum_{i=1}^{m} p_i l_i - \sum_{i=1}^{m} p_i \log_D \frac{1}{p_i}$$

$$= \sum_i p_i \log \frac{1}{D^{-l_i}} - \sum_i p_i \log_D \frac{1}{p_i}$$

$$= \sum_i p_i \log \frac{\sum_i D^{-l_i}}{D^{-l_i}} - \sum_i p_i \log_D \frac{1}{p_i} + \sum_i p_i \log_D \frac{1}{\left(\sum_i D^{-l_i}\right)}$$

$$= \sum_i p_i \log \frac{p_i}{q_i} + \log_D \left(\frac{1}{\sum_i D^{-l_i}} \right)$$

$$= D(p \| q) + \log_D \left(\frac{1}{\sum_i D^{-l_i}} \right) \geq 0$$

Equality holds iff,

i) $\sum_i D^{-l_i} = 1$, and

ii) $p_i = q_i = D^{-l_i}$ (source must be D-adic)

$$\therefore \boxed{L \geq H_D(x)}$$

__Theorem-__

Let $l_1^* \ldots\ldots l_m^*$ be optimal codeword lengths for a source distribution p and a D-ary alphabet. Let

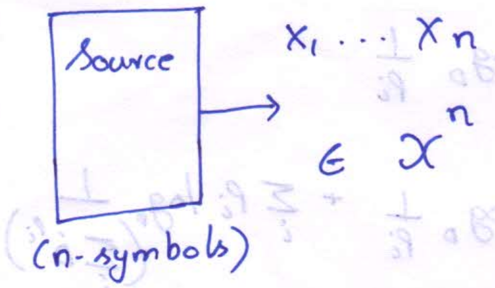
$$L^* = \sum_i p_i l_i^* \quad \left(\begin{array}{c}\text{expected}\\ \text{length}\end{array}\right)$$

Then

$$H_D(x) \leq L^* < H_D(x) + 1$$

__Proof-__
we have established the left inequality, the inequality on right follows from construction, $l_i = \left\lceil \log_D \frac{1}{p_i} \right\rceil$

Next Consider



Associated distribution $p(x_1, \ldots, x_n)$
Then by following what we have just done with single symbol replaced by strings, we will obtain,

$$\frac{H_D(x_1, \ldots, x_n)}{n} \leq L_n < \frac{H_D(x_1, \ldots, x_n)}{n} + \frac{1}{n}$$

where,

$$L_n = \frac{\left(\sum_{\substack{x_1 \ldots x_n \\ \in \mathcal{X}^n}} p(x_1 \ldots x_n)\, l(x_1, \ldots x_n)\right)}{n}$$

Theorem -

The minimum expected codeword length per symbol satisfies -

$$\frac{H_D(x_1 \cdots x_n)}{n} \leq L_n^* < \frac{H_D(x_1 \cdots x_n)}{n} + \frac{1}{n}$$

Moreover, if $\{X_n\}$ is a stationary Random sequence,

$$L_n^* \longrightarrow H_D(x_0)$$ the entropy rate.

x———x

Theorem - (Wrong Code)

The expected length under $p(x)$ of the code length assignment:

$$l(x) = \left\lceil \log \frac{1}{q(x)} \right\rceil$$ satisfies

$$H(p) + D(p\|q) \leq E_p[l(x)] \nleq H(p) + D(p\|q) + 1$$

Proof -

Consider,

$$E_p[l(x)] = \sum_x p(x) \left\lceil \log \frac{1}{q(x)} \right\rceil$$

$$< \sum_x p(x) \left(\log \frac{1}{q(x)} \right) + 1$$

$$= \sum_x p(x) \log \frac{p(x)}{q(x)} + \sum_x p(x) \log \frac{1}{p(x)} + 1$$

$$= D(p\|q) + H(p) + 1$$

(the other inequality can be similarly proved)

x———x