

# Math 5251 Kraft & McMillan inequalities (§3.2)

Can we **arbitrarily** specify the word lengths  $l_1, \dots, l_m$  for a code  $C = \{w_1, \dots, w_m\}$  on alphabet  $\Sigma$  of size  $n$ ?

called an  **$n$ -ary** alphabet/code  
e.g.  $\Sigma = \{0,1\}$  binary = 2-ary  
 $\Sigma = \{0,1,2\}$  ternary = 3-ary

**EXAMPLE**  $C = \{0, 1, 20, 21, 22\}$  on  $\Sigma = \{0,1,2\}$   
has  $(l_1, l_2, l_3, l_4, l_5)$   $m=5$   $n=3$   
 $= (1, 1, 2, 2, 2)$

Certainly **not arbitrarily**, e.g. if  $\Sigma = \{0,1\}$

then  $(l_1, l_2, l_3, l_4, l_5) = (2, 2, 2, 2, 2)$

is **impossible** since  $\Sigma^*$  has only

4 words of length 2:

	00
	01
	10
	11

If we further insist on the code being **uniquely decipherable**, it imposes even more of a constraint on  $(l_1, \dots, l_m)$ ; interestingly it's the same constraint for codes that are prefix.

---

**THEOREM** Let  $\Sigma$  be an alphabet with  $n$  letters, and  $(l_1, l_2, \dots, l_m)$  positive integers.

(a) (**Kraft**) If  $\sum_{i=1}^m \frac{1}{n^{l_i}} = \frac{1}{n^{l_1}} + \frac{1}{n^{l_2}} + \dots + \frac{1}{n^{l_m}} \leq 1$

then  $\exists$  a **prefix code**  $\mathcal{C}$  on  $\Sigma$  with those lengths.  
(instantaneous)

(b) (**McMillan**) If  $\exists$  a **uniquely decipherable** code  $\mathcal{C}$  on  $\Sigma$  with those lengths,

then  $\sum_{i=1}^m \frac{1}{n^{l_i}} \leq 1$

---

**SAME inequality for both!** So one concludes

{lengths of **u.d.**  $n$ -ary codes}

$\equiv$

{lengths of **prefix**  $n$ -ary codes}

$\equiv$  {  $(l_1, \dots, l_m)$  with  $\sum_i \frac{1}{n^{l_i}} \leq 1$  }  $\equiv$

*Kraft-McMillan inequality*

EXAMPLES If  $n=3=|\Sigma|$ , say  $\Sigma=\{0,1,2\}$

then  $\nexists$  any u.d. code  $\mathcal{C}$  with word lengths  $(1,1,2,2,2,3)$  because

$$\frac{1}{3^1} + \frac{1}{3^1} + \frac{1}{3^2} + \frac{1}{3^2} + \frac{1}{3^2} + \frac{1}{3^3} = \frac{9+9+3+3+3+1}{27} = \frac{28}{27} > 1$$

On the other hand, there does  $\exists$  a prefix code  $\mathcal{C}$  with lengths  $(1,2,2,2,3,3)$  because

$$\frac{1}{3^1} + \frac{1}{3^2} + \frac{1}{3^2} + \frac{1}{3^2} + \frac{1}{3^3} + \frac{1}{3^3} = \frac{9+3+3+3+3+1+1}{27} = \frac{23}{27} \leq 1$$

---

In fact, let's prove Kraft first, via an algorithm to find  $\mathcal{C}$ . Assuming  $(l_1, \dots, l_m)$  has  $t_i$  occurrences of length  $i$ , then the inequality assumes

$$\sum_{i=1}^m \frac{1}{n^{li}} = \frac{t_1}{n^1} + \frac{t_2}{n^2} + \frac{t_3}{n^3} + \dots \leq 1$$

and we try to pick the shorter words first.

EXAMPLE  $(l_1, \dots, l_m) = (1, 2, 2, 2, 2, 3, 3)$

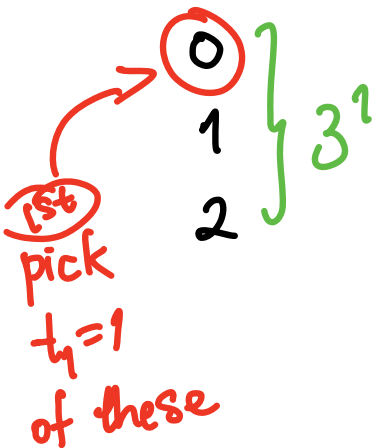
$\underbrace{1}_{t_1=1} \quad \underbrace{2, 2, 2, 2}_{t_2=4} \quad \underbrace{3, 3}_{t_3=2}$

has  $\frac{t_1}{3^1} + \frac{t_2}{3^2} + \frac{t_3}{3^3} = \frac{1}{3^1} + \frac{4}{3^2} + \frac{2}{3^3} \leq 1$

$\Rightarrow \frac{t_1}{3^1} \leq 1$

so  $t_1 \leq 3^1$

allowing us to pick  $t_1$  words of length 1:



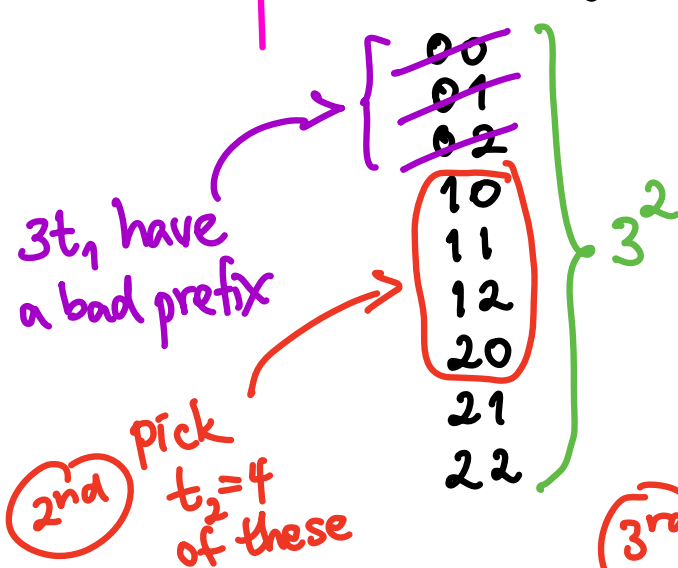
$\Rightarrow \frac{t_1}{3^1} + \frac{t_2}{3^2} \leq 1$

$\Leftrightarrow 3t_1 + t_2 \leq 3^2$

$t_2 \leq 3^2 - 3t_1$

this many length 2 words have a prefix from our length 1 choices

allowing us to pick  $t_2$  words of length 2:



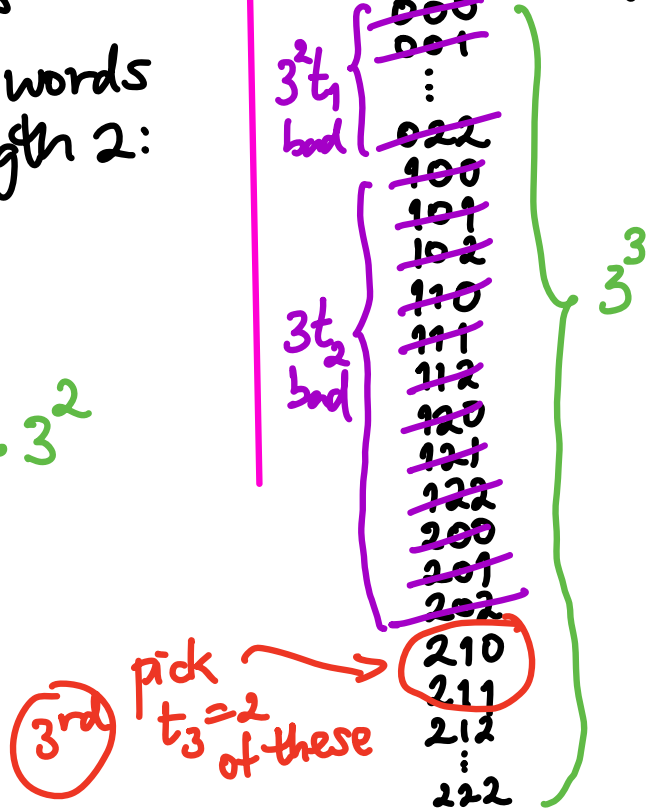
$\Rightarrow \frac{t_1}{3^1} + \frac{t_2}{3^2} + \frac{t_3}{3^3} \leq 1$

$\Leftrightarrow 3^2 t_1 + 3 t_2 + t_3 \leq 3^3$

$t_3 \leq 3^3 - (3^2 t_1 + 3 t_2)$

this many length 3 words have a bad prefix

allowing us to pick  $t_3$  words of length 3



## proof of Kraft's inequality:

If  $(l_1, \dots, l_m)$  has  $t_i$  occurrences of  $i$  and

$$\frac{t_1}{n^1} + \frac{t_2}{n^2} + \frac{t_3}{n^3} + \dots = \sum_{i=1}^m \frac{1}{n^{l_i}} \leq 1$$

we show how to pick a prefix code  $\mathcal{C}$  with those lengths. Assuming one has already picked the words of length  $\leq i-1$ , and show they leave  $\geq t_i$  words of length  $i$  that avoid them as prefixes.

Previously one has picked

$t_{i-1}$  of length  $i-1$   $\rightsquigarrow$  create  $n t_{i-1}$  with bad prefix

$t_{i-2}$  of length  $i-2$   $\rightsquigarrow$  create  $n^2 t_{i-2}$  with bad prefix

$\vdots$

$t_2$  of length 2  $\rightsquigarrow$  create  $n^{i-2} t_2$  with bad prefix

$t_1$  of length 1  $\rightsquigarrow$  create  $n^{i-1} t_1$  with bad prefix

Since there are  $n^i$  words of length  $i$

in total using alphabet  $\Sigma$ , ...

this leaves

$$n^i - (n^{i-1}t_1 + n^{i-2}t_2 + \dots + n^2t_{i-2} + nt_{i-1})$$

words of length  $i$  from which to choose  $t_i$  for  $\mathcal{C}$ .

We claim the above quantity is **at least**  $t_i$ ,

since 
$$\frac{t_1}{n^1} + \frac{t_2}{n^2} + \dots + \frac{t_{i-2}}{n^{i-2}} + \frac{t_{i-1}}{n^{i-1}} + \frac{t_i}{n^i} \leq 1$$

$\} \text{ multiply by } n^i$

$$n^{i-1}t_1 + n^{i-2}t_2 + \dots + n^2t_{i-2} + nt_{i-1} + t_i \leq n^i$$

i.e.  $t_i \leq n^i - (n^{i-1}t_1 + n^{i-2}t_2 + \dots + n^2t_{i-2} + nt_{i-1})$



# proof of McMillan inequality:

Assume  $\mathcal{C}$  is a uniquely decipherable  $n$ -ary code having  $t_i$  codewords of length  $i$  for  $i=1, 2, \dots, l$ .

We want to show  $\frac{t_1}{n^1} + \frac{t_2}{n^2} + \dots + \frac{t_l}{n^l} \leq 1$

call this sum  $A$ ; want  $A \leq 1$ .

**IDEA:** Instead, for each  $p=1, 2, 3, \dots$  we will show

$$A^p = \sum_{s=1}^{pl} \frac{c_s}{n^s} \text{ for some coefficients } c_s \leq n^s$$

$$\Rightarrow A^p \leq \sum_{s=1}^{pl} 1 = pl$$

$$\Rightarrow A \leq (pl)^{\frac{1}{p}}$$

take  $p^{\text{th}}$  root of both sides

$$\Rightarrow A \leq \lim_{p \rightarrow \infty} (pl)^{\frac{1}{p}} = 1, \text{ as desired}$$

$$\lim_{p \rightarrow \infty} (pl)^{\frac{1}{p}} = \lim_{p \rightarrow \infty} e^{\frac{\ln(pl)}{p}} = e^{\lim_{p \rightarrow \infty} \frac{\ln(p)}{p} + \frac{\ln(l)}{p}}$$

$$\stackrel{\text{L'Hôpital}}{=} e^{\lim_{p \rightarrow \infty} \frac{1/p}{1} + 0}$$

$$= e^0 = 1$$

Calculus  
interlude!

So for  $\mathcal{C}$  u.d., we need to show

$$A := \frac{t_1}{n^1} + \frac{t_2}{n^2} + \dots + \frac{t_l}{n^l} \text{ has } A^p = \sum_{s=1}^{pl} \frac{c_s}{n^s} \text{ with } c_s \leq n^s$$

In fact, we can interpret  $c_s$  as counting the number of messages  $(w_1, w_2, \dots, w_p)$  of  $p$  words from  $\mathcal{C}$  with a total length of  $s$  letters from  $\Sigma$ .

Since there are  $n^s$  strings in  $\Sigma^*$  with  $s$  letters, and  $\mathcal{C}$  is uniquely decipherable, this shows  $c_s \leq n^s$ ; each string comes from at most one message.

## ACTIVE LEARNING

If  $W \xrightarrow{f} \{0,1,2\}^*$  has codewords  $C = m(f) = \left\{ \begin{array}{l} t_1=1 \\ \text{of length 1} \end{array} \middle| \begin{array}{l} t_2=3 \\ \text{of length 2} \end{array} \right\} \{0,1, 20, 21, 22\}$ ,

(a) how many messages  $(w_1, w_2)$  with 2 source words end up with encodings  $f^*(w_1, w_2) = f^*(w_1)f^*(w_2)$  that are

2 letters long (like 1|0 or 0|1) ?

3 letters long (like 1|21 or 22|0) ?

4 letters long ?

5 letters long ?

(b) How would those answers change if  $\Sigma$  had more letters and

$t_1 = 10$

$t_2 = 1000$  ?



(proof by)  
**EXAMPLE** that  $A := \frac{t_1}{n^1} + \frac{t_2}{n^2} + \dots + \frac{t_l}{n^l}$

has  $A^p = \sum_{s=1}^{pl} \frac{C_s}{n^s}$  with  $C_s \leq n^s$

where

$C_s =$  number of messages  $(w_1, w_2, \dots, w_p)$  of  $p$  words from  $\mathcal{C}$  with a total length of  $s$  letters from  $\Sigma$ .

**EXAMPLE:**  $\mathcal{C} = \{0, 1, 20, 21, 22\}$ ,  $\Sigma = \{0, 1, 2\}$   
 $n=3$   
 $t_1=2$   $t_2=3$

$$\left(\frac{t_1}{3^1} + \frac{t_2}{3^2}\right)^2 = \frac{t_1 \cdot t_1}{3^2} + \frac{(t_1 t_2 + t_2 t_1)}{3^3} + \frac{t_2 \cdot t_2}{3^4}$$

$$= \frac{2 \cdot 2}{3^2} + \frac{2 \cdot 3 + 3 \cdot 2}{3^3} + \frac{3 \cdot 3}{3^4}$$

00  
01  
10  
11

020 200  
021 201  
022 210  
120 211  
121 220  
122 221

2020  
2021  
2022  
2120  
2121  
2122  
2220  
2221  
2222



RECAP: We showed

(u.d.  $\Leftrightarrow$  prefix)

$\exists$  a u.d.  $n$ -ary code  $\mathcal{C}$  with lengths  $(l_1, \dots, l_m)$



$\exists$  a prefix  $n$ -ary code  $\mathcal{C}$  with lengths  $(l_1, \dots, l_m)$

McMillan



Kraft

$$\sum_{i=1}^m \frac{1}{n^{l_i}} \leq 1$$

so all 3 statements are equivalent.