University of Windsor

Scholarship at UWindsor

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

2022

The Equality Case of the Kraft and the Kraft-McMillan Inequalities

Xavier Nunes
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Part of the Mathematics Commons

Recommended Citation

Nunes, Xavier, "The Equality Case of the Kraft and the Kraft-McMillan Inequalities" (2022). *Electronic Theses and Dissertations*. 9008.
https://scholar.uwindsor.ca/etd/9008

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

The Equality Case of the Kraft and the Kraft-McMillan Inequalities

by

Xavier Nunes

A Thesis

Submitted to the Faculty of Graduate Studies

through the Department of Mathematics and Statistics

in Partial Fulfillment of the Requirements for

the Degree of Master of Science at the

University of Windsor

Windsor, Ontario, Canada

© 2022 Xavier Nunes

The Equality Case of the Kraft and the Kraft-McMillan Inequalities

by

Xavier Nunes

APPROVED BY:

A. Mukhopadhyay

School of Computer Science

I. Shapiro

Department of Mathematics and Statistics

A. Zahariuc, Advisor

Department of Mathematics and Statistics

August 23, 2022

Author's Declaration or Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

In this thesis, we analyze the Kraft Inequality and the Kraft-McMillan Inequality in their equality cases. Kraft's Inequality deals with prefix-free code and Kraft-McMillan's Inequality deals with uniquely decodable codes. The focus of the Kraft Inequality analysis is to study the occurrence of prefix-free codes that satisfy the equality case and the structure of words in the code when the equality condition is met. The second part of the thesis touches on the Kraft-McMillan Inequality. Since the proof of this latter inequality uses limits, we cannot immediately analyse its equality cases. The paper will therefore study the equality cases of this theorem and demonstrate that these equality cases have similar results to that of the Kraft Inequality, although it is necessary to prove them in a different way since the latter theorem's proof is less direct.

Acknowledgements

I would first like to thank my advisor, Dr. Zahariuc for his incredible support during the course of my master's degree. Without his guidance I would not have been able to complete this project in the way that I have without his guidance. Dr. Zahariuc's knowledge and experience were able to guide to achieve this milestone. A special thanks is extended to Dr. Shapiro from the Department of Mathematics and Statistics and Dr. Asish Mukhopadhyay from the School of Computer Science for offering their time in reading my thesis and for being a part of my defence committee. I am also grateful to Dr. Dennis Borisov for chairing my defence.

Next, I would like to thank my professors, my fellow students and faculty members to have always offered their support when needed. I would not have been able to complete my courses let alone my final thesis paper without their support and guidance.

Finally, I would like to offer my greatest gratitude to my parents and brothers for their unfailing support and for pushing me from front to end in the process of my education, words cannot express how much their support has meant to me. My grandparents as well as aunts have show me their love and support at every step of the way.

Contents

Chapter 1

Introduction

This thesis is concerned with two theorems in Information Theory: the Kraft Inequality and the Kraft-McMillan Inequality. These theorems deal with prefix-free codes and uniquely decodable codes respectively. We start by reviewing these notions and the context in which they occur.

1.1 Basic Notation

We start by introducing the language that will be used in this thesis. An *alphabet* $\mathcal{A}$ is a finite set of *characters* (which we sometimes call *letters*). Let $d = |\mathcal{A}|$, so there are d letters in the alphabet. Let $\mathcal{A}^*$ be the free monoid generated by $\mathcal{A}$. In computer science literature, $\mathcal{A}^*$ is often called the Kleene star of $\mathcal{A}$. A *word* is an element of $\mathcal{A}^*$ and thus is a string of characters from $\mathcal{A}$.

A *code* is a subset $\mathcal{C}$ of $\mathcal{A}^*$. We will only consider finite codes $\mathcal{C} = \{w_1, \ldots, w_n\}$ where each w_i is a word in $\mathcal{A}^*$. For any $w \in \mathcal{C}$, let $l(w)$ be its length.

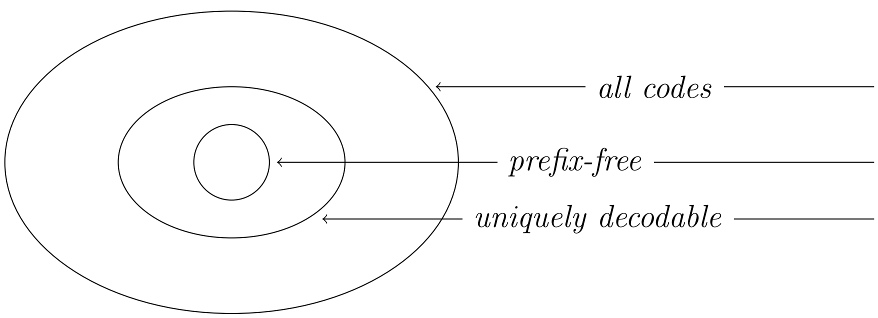
1.2 The Kraft and Kraft-McMillan Inequalities

Definition 1.2.1. The notation l_{max} is used to represent the longest word in a code $\mathcal{C}$, i.e. $l_{max} = \max\{l(w_1), \ldots, l(w_n)\}$.

Definition 1.2.2. A code $\mathcal{C}$ is *prefix-free* if there do not exist two distinct words $w, v \in \mathcal{C}$ such that w is an initial segment of v. Similarly, a code is *suffix-free* if there do not exist two distinct words $w, v \in \mathcal{C}$ such that w is an ending segment of v.

Definition 1.2.3. A code $\mathcal{C}$ is *uniquely decodable* if any word in $\mathcal{A}^*$ can be written as a concatenation of words in $\mathcal{C}$ in at most one way.

Clearly, any prefix-free code is uniquely decodable.



However, the converse is not true: there exist uniquely decodable codes which are not prefix-free.

Example 1.2.4. If we take an alphabet $\mathcal{A}$ and a code $\mathcal{C}$ to be suffix-free, this gives a simple way of getting a uniquely decodable code that is not prefix-free. For example, if $\mathcal{A} = \{a, b\}$, and the suffix-free code is $\mathcal{C} = \{a, ab, abb, bbb\}$, then this code is clearly uniquely decodable because it is suffix-free and it is not prefix-free.

Example 1.2.5. Let $\mathcal{A} = \{0, 1\}$ be our alphabet and $\mathcal{C} = \{0, 001, 110\}$ be our code. Clearly, $\mathcal{C}$ is not prefix-free, as 0 is a prefix to 001. However, it is uniquely decodable because of the conditions in which we can find a 1. If we can find an isolated 1, then the word associated must be *001*. If there are two 1's together we must have *110* and if there are 3 one's in a row then the result must be *001* followed by *110*. Now outside of these occasions the rest of the characters will be 0's from the word *0*.

Now we can state the Kraft and the Kraft-McMillan Inequalities.

Theorem 1.2.6 (The Kraft Inequality)**.** *If $\mathcal{C}$ is a prefix-free code from the alphabet $\mathcal{A}$, then*

$$\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} \leq 1.$$

Proof. Let $C(w) = \{v \in \mathcal{A}^* : l(v) = l_{max}, w \text{ is a prefix of } v\}$. Since $\mathcal{C}$ is a prefix-free code, $C(w_1) \cap C(w_2) = \emptyset$ for all $w_1, w_2 \in \mathcal{C}$ such that $w_1 \neq w_2$. Then,

$$\sum_{w\in\mathcal{C}} |C(w)| \leq |\{v \in \mathcal{A}^* : l(v) = l_{max}\}|.$$

Hence,

$$\sum_{w\in\mathcal{C}} d^{l_{max}-l(w)} \leq d^{l_{max}}.$$

The inequality is now a straightforward result. ■

Theorem 1.2.7 (Kraft-McMillan Inequality)**.** *If $\mathcal{C}$ is a uniquely decodable code, then*

$$\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} \leq 1.$$

Proof. Let $m \in \mathbb{N}$ and $N_{m,l}$ be the number of combinations of m codewords with length l.

Since $\mathcal{C}$ is uniquely decodable, $|N_{m,l}| \leq d^l$. Then,

$$\begin{aligned}\left(\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}}\right)^m &= \sum_{w_1,w_2,\ldots,w_m\in\mathcal{C}} \frac{1}{d^{l(w_1)+l(w_2)+\ldots+l(w_m)}} \\ &= \sum_{l=0}^{m\cdot l_{max}} \frac{N_{m,l}}{d^l} \\ &\leq \sum_{l=0}^{m\cdot l_{max}} 1 \\ &= m\cdot l_{max}+1.\end{aligned}$$

Hence $\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} \leq \sqrt[m]{m\cdot l_{max}+1}$, but $\lim_{m\to\infty} \sqrt[m]{m\cdot l_{\max}+1} = 1$, so we get $\sum_{m\in\mathcal{C}} \frac{1}{d^{l(w)}} \leq 1$. ■

This proof of McMillan's first inequality [2] is the method used by Karush [3].

The second part of McMillan's work is the Shannon-McMillan Theorem. Shannon coding, the base of this part of his work, is a technique of lossless data compression for constructing a prefix-free code with their probabilities.

The Kraft-McMillan Inequality extends what we know from the Kraft Inequality for prefix-free codes to the more general uniquely decodable codes. Hence, when it satisfies the equality case, our code is full and cannot be enlarged.

Example 1.2.8. Here is a list of examples where equality is attained in the Kraft and Kraft-McMillan Inequalities.

1. If $\mathcal{A} = \{a\}$, then $\mathcal{C} = \{a\}$ is clearly uniquely decodable, and

$$\frac{1}{1^{l_a}} = 1.$$

2. If $\mathcal{A} = \{0,1\}$, then $\mathcal{C} = \{0, 11, 101, 100\}$ is clearly prefix-free, and

$$\sum_{w\in\mathcal{C}} \frac{1}{2^{l(w)}} = \frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{8} = 1.$$

3. If $\mathcal{A} = \{a, b, c\}$, then $\mathcal{C} = \{a, ba, bb, bc, c\}$ is prefix-free and

$$\sum_{w \in \mathcal{C}} \frac{1}{3^{l(w)}} = \frac{1}{3} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{3} = 1.$$

4. If $\mathcal{A} = \{0, 1\}$, then $\mathcal{C} = \{00, 01, 11, 101, 1000, 1001\}$ is prefix-free, and

$$\sum_{w \in \mathcal{C}} \frac{1}{2^{l(w)}} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = 1.$$

5. Let $\mathcal{C}$ be the prefix-free code that contains all words of length N for a given alphabet of size d. Then there are d^N words and we have

$$\sum_{w \in \mathcal{C}} \frac{1}{d^{l(w)}} = \sum_{w \in \mathcal{C}} \frac{1}{d^N} = d^N \frac{1}{d^N} = 1.$$

6. If $\alpha \in \mathcal{A}$ is fixed, let $\mathcal{C}$ consist of all words of the form

$$\underbrace{\alpha\alpha\ldots\alpha}_{\text{k occurrences}}\ \beta$$

with $0 \leq k \leq N-1$ and $\beta \in \mathcal{A}$ arbitrary such that $\beta \neq \alpha$ if $k \neq N-1$. For example, if

$$\mathcal{A} = \{a, b, c\}, \qquad \text{then } \mathcal{C} = \{b, c, ab, ac, \ldots, a\ldots ab, a\ldots ac, a\ldots aa\}.$$

Clearly $\mathcal{C}$ is prefix-free and

$$\sum_{w \in \mathcal{C}} \frac{1}{d^{l(w)}} = (d-1)\frac{1}{d} + (d-1)\frac{1}{d^2} + \ldots + d\frac{1}{d^N} = 1.$$

7. If $\mathcal{A} = \{0, 1\}$ and $\mathcal{C} = \{0, 01, 11\}$, then we see that $\mathcal{C}$ is clearly not prefix-free but is

suffix-free, so it is uniquely decodable. We then have

$$\sum_{w\in\mathcal{C}} \frac{1}{2^{l(w)}} = \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1.$$

1.3 The Results

Question 1.3.1. What can be said about $\mathcal{C}$ (a prefix-free or uniquely decodable code) if equality occurs in the Kraft Inequality or the Kraft-McMillan Inequality, i.e. if

$$\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} = 1?$$

Two more notions we have to define for our results in this thesis are the *occurrence* and the *weighted number of occurrences* of a character. The occurrence of a character, α, in a word $w \in \mathcal{C}$ is the number of times α appears in the word and we will use the notation $n_\alpha(w)$. The weighted number of occurrences is calculated for α in the entire code $\mathcal{C}$:

$$A_\alpha := \sum_{w\in\mathcal{C}} \frac{n_\alpha(w)}{d^{l(w)}}.$$

Example 1.3.2. We take the same examples as in Example 1.2.8 and calculate A_α for all $\alpha, \beta \in \mathcal{A}$. Each example correlates to the same number from the previous set.

1. $\mathcal{A} = \{a\}$ and $\mathcal{C} = \{a\}$, so
$$A_a = 1.$$

2. $\mathcal{A} = \{0, 1\}$ and $\mathcal{C} = \{0, 11, 101, 100\}$, so
$$A_0 = \frac{1}{2} + \frac{1}{8} + \frac{2}{8} = \frac{7}{8}; \quad A_1 = \frac{2}{4} + \frac{2}{8} + \frac{1}{8} = \frac{7}{8}.$$

3. $\mathcal{A} = \{a, b, c\}$ and $\mathcal{C} = \{a, ba, bb, bc, c\}$, so

$$A_a = \frac{1}{3} + \frac{1}{9} = \frac{4}{9}; \quad A_b = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}; \quad A_c = \frac{1}{9} + \frac{1}{3} = \frac{4}{9}.$$

4. $\mathcal{A} = \{0, 1\}$ and $\mathcal{C} = \{00, 01, 11, 101, 1000, 1001\}$, so

$$A_0 = \frac{2}{4} + \frac{1}{4} + \frac{1}{8} + \frac{3}{16} + \frac{2}{16} = \frac{19}{16}; \quad A_1 = \frac{1}{4} + \frac{2}{4} + \frac{2}{8} + \frac{1}{16} + \frac{2}{16} = \frac{19}{16}.$$

5. Let $\mathcal{C}$ be the set of all words of length N. Therefore, there are d^N words. There are Nd^N letters overall, and α will occur $\frac{1}{d}$ of the time. From this, we get $\sum_{w \in \mathcal{C}} n_\alpha(w) = \frac{N}{d} d^N$, and so for all $\alpha \in \mathcal{A}$,

$$A_\alpha = \frac{1}{d^N} \sum_{w \in \mathcal{C}} n_\alpha(w) = N d^{N-1} \frac{1}{d^N} = \frac{N}{d}.$$

6. In the two sections in this example we will use

$$\sum_{k=0}^{n} k r^k = r \frac{1 - (n+1) r^{n-1} + n r^{n+1}}{(1-r)^2}.$$

Indeed, multiplying both sides by $(1-r)^2$ and multiplying out on the left hand side boils this down to a routine calculation. This calculation can be seen at [8].

(a) $\mathcal{A} = \{a, b, c\}$ and $\mathcal{C} = \{b, c, ab, ac, \ldots, a \ldots ab, a \ldots ac, a \ldots aa\}$, so

$$A_b = A_c = \frac{1}{3} + \frac{1}{3^2} + \ldots + \frac{1}{3^{N-1}} = \frac{1}{2}\left(1 - \frac{1}{3^N}\right)$$

$$\begin{aligned} A_a &= \frac{2}{3^2} + \frac{2\cdot 2}{3^3} + \ldots + \frac{2\cdot(N-1)}{3^N} + \frac{N}{3^N} \\ &= \frac{2}{3}\left(\frac{1}{3} + \frac{2}{3^2} + \frac{3}{3^3} + \ldots + \frac{N-1}{3^{N-1}}\right) + \frac{N}{3^N} \\ &= \frac{2}{3}\left(\left(\frac{1}{3}\right)\frac{1 - N\frac{1}{3^{N-1}} + (N-1)\frac{1}{3^N}}{(1-\frac{1}{3})^2}\right) + \frac{N}{3^N} \\ &= \frac{1}{2}\left(1 - N\frac{1}{3^{N-1}} + (N-1)\frac{1}{3^N} + \frac{2N}{3^N}\right) \\ &= \frac{1}{2}\left(1 - \frac{1}{3^N}\right). \end{aligned}$$

(b) For a fixed α and an alphabet of size d, we have

$$\begin{aligned} A_\alpha &= \frac{d-1}{d^2} + \frac{2(d-1)}{d^3} + \ldots + \frac{(d-1)(N-1)}{d^N} + \frac{N}{d^N} \\ &= \frac{(d-1)}{d}\left(\frac{1}{d} + \frac{2}{d^2} + \ldots + \frac{N-1}{d^{N-1}}\right) + \frac{N}{d^N} \\ &= \frac{d-1}{d}\left(\frac{1}{d}\cdot\frac{1 - N\frac{1}{d^{N-1}} + (N-1)\frac{1}{d^N}}{(1-\frac{1}{d})^2}\right) + \frac{N}{d^N} \\ &= \frac{1}{d-1}\left(1 - N\frac{1}{d^{N-1}} + (N-1)\frac{1}{d^N} + \frac{(d-1)N}{d^N}\right) \\ &= \frac{1}{d-1}\left(1 - \frac{1}{d^N}\right) \end{aligned}$$

$$A_\beta = \frac{1}{d} + \frac{1}{d^2} + \ldots + \frac{1}{d^{N-1}} = \frac{1}{d}\left(\frac{1-\frac{1}{d^N}}{1-\frac{1}{d}}\right) = \frac{1}{d-1}\left(1 - \frac{1}{d^N}\right).$$

7. $\mathcal{A} = \{0, 1\}$ and $\mathcal{C} = \{0, 01, 11\}$, so

$$A_0 = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}; \quad A_1 = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}.$$

Theorem 1.3.3. *Let $\mathcal{C}$ be prefix-free and $\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} = 1$. (I.e. the equality case occurs in Kraft's Inequality.) Then,*

$$A_\alpha = A_\beta \quad \text{for all } \alpha, \beta \in \mathcal{A}. \tag{1.1}$$

The next theorems we go over are all refinements on this base theorem.

Theorem 1.3.4. *Let $\mathcal{C}$ be a prefix-free code. We have*

$$\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} \leq 1 - \frac{1}{l_{max}} \max_{\alpha,\beta\in\mathcal{A}} \{|A_\alpha - A_\beta|\}. \tag{1.2}$$

In the above theorem, we find a sharpened version of the Kraft Inequality from which we get Theorem 1.3.3 when equality occurs in the Kraft Inequality and the right hand side is strictly 1.

Theorem 1.3.5. *If $\mathcal{C}$ is a prefix-free code that satisfies the equality case of the Kraft Inequality, then*

1. $A_\alpha = A_\beta$
2. $\displaystyle\sum_{w\in\mathcal{C}} \frac{(n_\alpha(w) - n_\beta(w))^2}{d^{l(w)}} = A_\alpha + A_\beta$

for all $\alpha, \beta \in \mathcal{A}$.

The above theorem is an extension of Theorem 1.3.3 but offers more information regarding the weighted number of occurrences of characters.

Theorem 1.3.6. *Assume that $\mathcal{C}$ is uniquely decodable and that $\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} = 1$ (i.e. that equality occurs in the Kraft-McMillan Inequality). Then*

$$A_\alpha = A_\beta, \quad \text{for all } \alpha, \beta \in \mathcal{A}.$$

Since the proof of the Kraft-McMillan Inequality uses limits, this proof says very little about when the equality cases of the theorem are satisfied. To prove Theorem 1.3.6 we will rely on a generalization of the Kraft-McMillan Inequality (Proposition 4.1.1) and combine it with a variational approach using Lagrange multipliers. This raises the question about whether the behaviour in the equality case of the Kraft Inequality is actually similar to the behaviour of the equality case in the Kraft-McMillan Inequality as well.

We will arrive to the result of Theorem 1.3.3 in three different ways in this thesis and each of these results offers different additional results.

Chapter 2

Preliminaries

2.1 Reviewing Monoids

Throughout this thesis we will use monoids and monoid homomorphisms, mainly as a convenient language device.

Definition 2.1.1 (Monoid)**.** A *monoid* is a triple $(M, e, *)$ consisting of a set M, an associative binary operation $*$ on M and an identity element e. In other words, monoids satisfy the following rules:

1. $(x * y) * z = x * (y * z)$ for all $x, y, z \in M$
2. $e * x = x * e = x$ for all $x \in M$.

Definition 2.1.2. A *monoid homomorphism* between two monoids $(M_1, e_1, *)$ and $(M_2, e_2, \circ)$ is a function $\phi : M_1 \to M_2$ such that:

1. $\phi(x * y) = \phi(x) \circ \phi(y)$ for all $x, y \in M_1$
2. $\phi(e_1) = e_2$.

Remark 2.1.3. If $\phi : (M_1, e_1, *) \to (M_2, e_2, \circ)$ is a monoid homomorphism, then

$$\phi(x_1 * x_2 * \ldots * x_n) = \phi(x_1) \circ \phi(x_2) \circ \ldots \circ \phi(x_n) \quad \text{for all } x_1, x_2, \ldots, x_n \in M_1 \tag{2.1}$$

(i.e. the property in the definition remains true for any number of elements).

2.2 Important Notions

First recall that $\mathcal{A}^*$ is the free monoid generated by $\mathcal{A}$ and we will simply use this notation instead of $(\mathcal{A}^*, " \ ", *)$. We will work with the monoid homomorphism from $\mathcal{A}^*$ to $[0, \infty)$ with the latter under multiplication. These have a very specific form, i.e.

$$\phi(v) = \prod_{\alpha \in \mathcal{A}} x_\alpha^{n_\alpha(v)} \tag{2.2}$$

where $x_\alpha \geq 0$ is a constant (in fact, $x_\alpha = \phi(\alpha)$).

Lemma 2.2.1. *Let* $\phi : \mathcal{A}^* \to [0, \infty)$ *and* $L \in \mathbb{N}$*. Then,*

$$\sum_{\substack{v \in \mathcal{A}^* \\ l(v) = L}} \phi(v) = \left(\sum_{\alpha \in \mathcal{A}} \phi(\alpha) \right)^L .$$

Proof.

$$\begin{aligned} \sum_{\substack{v \in \mathcal{A}^* \\ l(v) = L}} \phi(v) &= \sum_{\substack{v \in \mathcal{A}^* \\ l(v) = L}} \prod_{\alpha \in \mathcal{A}} \phi(\alpha)^{n_\alpha(v)} \quad \text{by (2.2) and } x_\alpha = \phi(\alpha) \\ &= \sum_{\sum k_\alpha = L} \binom{L}{k_{\alpha_1}, \ldots, k_{\alpha_d}} \prod_{\alpha \in \mathcal{A}} \phi(\alpha)^{k_\alpha} \\ &= \left(\sum_{\alpha \in \mathcal{A}} \phi(\alpha) \right)^L \quad \text{by the Multinomial Theorem.} \end{aligned}$$

■

We say that $\phi \in Hom(\mathcal{A}^*, [0, \infty))$ is *normalized* if $\sum_{\alpha \in \mathcal{A}} \phi(\alpha) = 1$. Let $\mathcal{H} \subset Hom(\mathcal{A}^*, [0, \infty))$ be the set of normalized homomorphisms. In the next chapters, we will often be using $\mathcal{H}^\circ = \{\phi \in \mathcal{H} : \phi(\alpha) > 0, \text{for all } \alpha \in \mathcal{A}\}$. Let

$$\Psi : \mathcal{H}^\circ \to [0, \infty) \quad \text{defined as} \quad \Psi(\phi) = \sum_{w \in \mathcal{C}} \phi(w). \tag{2.3}$$

In explicit calculations, it is convenient to use more explicit notation as follows. We choose $\alpha_1, \alpha_2, \ldots, \alpha_d$ to be an enumeration of $\mathcal{A}$, and write $\Psi(x_1, \ldots, x_d)$ instead of $\Psi(\phi)$ where $x_i = x_{\alpha_i} = \phi(\alpha_i)$. Thus we can identify $\mathcal{H}$ with the $(d-1)$-simplex

$$\Delta = \{(x_1, x_2, \ldots, x_d\} \in \mathbb{R}^d : x_i \geq 0, \sum x_i = 1\}. \tag{2.4}$$

Let also $\phi_0 \in \mathcal{H}$, given by:

$$\phi_0(w) = \frac{1}{d^{l(w)}}. \tag{2.5}$$

Here we see that $\Psi(\phi_0)$ is the quantity $\sum_{w \in \mathcal{C}} \frac{1}{d^{l(w)}}$ which is exactly the left hand side in the Kraft and Kraft McMillan Inequalities.

Chapter 3

The Equality case of the the Kraft Inequality

This chapter studies what happens when equality occurs in the Kraft Inequality.

3.1 A Refinement of the Kraft Inequality

Our first result is an improvement of Kraft's Inequality.

Theorem 3.1.1. *If $\mathcal{C} \subset \mathcal{A}^*$ is a (finite) prefix-free code, then*

$$\sum_{w \in \mathcal{C}} \frac{1}{d^{l(w)}} \leq 1 - \frac{1}{l_{max}} \max_{\alpha, \beta \in \mathcal{A}} \{|A_\alpha - A_\beta|\}. \tag{3.1}$$

Before proving this theorem, we first need some preliminaries.

Lemma 3.1.2. *In the situation of Theorem 3.1.1,*

$$d\mathcal{A}_\beta - \sum_{\alpha \in \mathcal{A}} A_\alpha \leq l_{max} \left(1 - \sum_{w \in \mathcal{C}} \frac{1}{d^{l(w)}}\right)$$

for all $\beta \in \mathcal{A}$.

Proof. For $w \in \mathcal{C}$ and $N = l_{max}$, let $C(w)$ be the set of all length N words which start with w. Therefore, the size of the set $C(w)$ will be $d^{N-l(w)}$ as we include all possible endings to this prefix. Next, we calculate the amount of occurrences of β in $C(w)$. We know that $n_\beta(w)$ is the amount of occurrences of β in the first $l(w)$ letters in each word, or the prefix, thus we just need to find the number of occurrences in the remaining $N - l(w)$ letters. Clearly, we will get exactly $\frac{1}{d}^{\text{th}}$ of the remaining letters to be β, which is $d^{N-l(w)}\dfrac{N-l(w)}{d}$ resulting in β occurring $d^{N-l(w)}\left(n_\beta(w) + \dfrac{N-l(w)}{d}\right)$ times in $C(w)$.
Therefore, we get

$$\sum_{w\in\mathcal{C}} d^{N-l(w)}\left(n_\beta(w) + \frac{N-l(w)}{d}\right) \leq N d^{N-1}.$$

By reorganizing and diving by d^N, we have

$$\sum_{w\in\mathcal{C}} \frac{n_\beta(w)}{d^{l(w)}} + \frac{N}{d}\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} - \sum_{w\in\mathcal{C}} \frac{l(w)}{d^{l(w)}} \leq \frac{N}{d}.$$

Thus, we get

$$dA_\beta - \sum_{\alpha\in\mathcal{A}} A_\alpha \leq N\left(1 - \sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}}\right).$$

■

Lemma 3.1.3. *Let $x_1, x_2, \ldots, x_n \geq 0$. Then,*

$$\max_{1\leq i,j\leq n} |x_i - x_j| \leq \max_{1\leq k\leq n}\left(n x_k - \sum_{i=1}^{n} x_i\right).$$

Proof. To begin, let $a, b \in \{1, 2, \ldots, n\}$ such that $x_a = \max\limits_{1\leq k\leq n} x_k$ and $x_b = \min\limits_{1\leq k\leq n} x_k$. Then we have

$$\begin{aligned}\max_{1\le k\le n}\left(nx_k - \sum_{i=1}^{n} x_i\right) &= nx_a - \sum_{i=1}^{n} x_i \\ &= x_a - x_b + (n-1)x_a - \sum_{i\neq b} x_i \\ &= x_a - x_b + \sum_{i\neq b}(x_a - x_i) \\ &\ge x_a - x_b \\ &= \max_{1\le i,j\le n} |x_i - x_j|\end{aligned}$$

■

Proof of Theorem 3.1.1. The proof of the theorem follows from Lemmas 3.1.2 and 3.1.3, the latter applied with A_α's instead of x_i's. ■

Corollary 3.1.4. *Let $\alpha, \beta \in \mathcal{A}$ and suppose equality occurs in the Kraft Inequality. Then,*

$$A_\alpha = A_\beta.$$

Proof. Suppose we have $\sum_{w\in\mathcal{C}} \frac{1}{d^{l(w)}} = 1$ (the equality case of Kraft's Inequality). Using Theorem 3.1.1, we know that $\frac{1}{l_{max}} \max_{\alpha,\beta\in\mathcal{A}}\{|A_\alpha - A_\beta|\} = 0$. Therefore, $A_\alpha = A_\beta$ for all $\alpha, \beta \in \mathcal{A}$.

■

This is in fact the base theorem that will also follow from two other theorems that will be proven later on in the thesis.

3.2 Maximal prefix-free codes

Definition 3.2.1. A prefix-free code $\mathcal{C}$ is *maximal* if for any word w' such that $w' \notin \mathcal{C}$, then $\mathcal{C} \cup \{w'\}$ is not prefix-free.

Proposition 3.2.2. *If $\phi \in \mathcal{H}^\circ$ and $\mathcal{C}$ is a prefix-free code, then*

$$\Psi(\phi) \leq 1.$$

Moreover, $\Psi(\phi) = 1$ if and only if $\mathcal{C}$ is maximal.

Proof. Take $C(w) = \{v \in \mathcal{A}^* : l(v) = l_{max}, w \text{ is a prefix of } v\}$. As in the proof of the usual Kraft Inequality, if $\mathcal{C}$ is prefix-free, then

$$C(w_1) \cap C(w_2) = \emptyset \quad \text{for all } w_1, w_2 \in \mathcal{C} \text{ where } w_1 \neq w_2. \tag{3.2}$$

Now,

$$\begin{aligned}
\Psi(\phi) &= \sum_{w \in \mathcal{C}} \phi(w) \\
&= \sum_{w \in \mathcal{C}} \phi(w) \sum_{l(u)=N-l(w)} \phi(u) \text{ by Lemma 2.2.1} \\
&= \sum_{w \in \mathcal{C}} \sum_{l(u)=N-l(w)} \phi(w * u) \\
&= \sum_{w \in \mathcal{C}} \sum_{v \in C(w)} \phi(v) \\
&\leq \sum_{\substack{v \in \mathcal{A}^* \\ l(v)=N}} \phi(v) \quad \text{by (3.2)} \\
&= 1 \quad \text{by Lemma 2.2.1.}
\end{aligned}$$

Next, we will show that $\Psi(\phi) = 1$ if and only if $\mathcal{C}$ is maximal. To get this last result, we need only show that $\mathcal{C}$ is maximal if and only if $\bigcup_{w \in \mathcal{C}} C(w)$ is equal to the set of all length N words (i.e. $\{C(w) : w \in \mathcal{C}\}$ is a partition). First, we show that $\mathcal{C}$ is maximal if and only if $\bigcup_{w \in \mathcal{C}} C(w)$ is the set of all length N words. Assume $\{C(w) : w \in \mathcal{C}\}$ is a partition and we take the length N of words in $C(w)$ to be greater or equal to the length of the word, w' we try to add to $\mathcal{C}$. Therefore, w' will be a prefix of words in $\bigcup_{w \in \mathcal{C}} C(w)$. This implies that w is a prefix of w' or vice versa and thus $\mathcal{C}$ is not maximal. Conversely, if $\{C(w) : w \in \mathcal{C}\}$ is not a partition, then there exists a word w' of length N such that $w' \notin C(w)$ for all $w \in \mathcal{C}$. Thus $\mathcal{C} \cup \{w'\}$ is prefix-free, so $\mathcal{C}$ is not maximal.

Second, we show that $\bigcup_{w \in \mathcal{C}} C(w)$ is the set of all length N words if and only if $\Psi(\phi) = 1$. This follows from the sequences of inequalities above and $\phi \in \mathcal{H}^{\circ}$. ■

Corollary 3.2.3. *If $\phi \in \mathcal{H}^{\circ}$ and $\mathcal{C}$ is prefix-free with*

$$\Psi(\phi) = 1,$$

then

$$\sum_{w \in \mathcal{C}} \frac{1}{d^{l(w)}} = 1.$$

Proof. From Proposition 3.2.2, we have that C must be maximal. Now, by applying Proposition 3.2.2 again with $\phi = \phi_0$ it follows that $\Psi(\phi_0) = \sum_{w \in \mathcal{C}} \frac{1}{d^{l(w)}} = 1$. ■

Lemma 3.2.4. *If $F \in \mathbb{R}[x_1, x_2, \ldots, x_d]$ is a polynomial such that $F(x_1, x_2, \ldots, x_d) = 1$ for all $x_1, x_2, \ldots, x_d$ such that $x_1 + x_2 + \ldots + x_d = 1$ and $x_1, \ldots, x_d > 0$, then*

1. $\frac{\partial F}{\partial x_i} = \frac{\partial F}{\partial x_j}$ *for all i, j at points where $x_1 + x_2 + \ldots + x_d = 1$*
2. $2\partial_{ij} F = \partial_i^2 F + \partial_j^2 F$ *for all i, j at points where $x_1 + x_2 + \ldots + x_d = 1$*

are both true.

Remark 3.2.5. In fact there are higher derivatives that can be calculated to give new equations. However, the results are restricted to the first and second as the rest will be calculated in similar manners.

Proof. To begin, $1 - x_1 - x_2 - \ldots - x_d$ is irreducible and $F(x_1, \ldots, x_d) - 1 = 0$ whenever $1 - x_1 - \ldots - x_d = 0$ and $x_1, \ldots, x_d > 0$. It follows that $F(x_1, \ldots, x_d) - 1 = 0$ whenever $1 - x_1 - \ldots - x_d = 0$ and $x_i \in \mathbb{C}$. This follows from that fact that if $P \in \mathbb{R}[x_1, \ldots, x_m]$ such that $P(x_1, \ldots, x_m) = 0$ for all $(x_1, \ldots, x_m) \in U$ where U is an open set, then $P = 0$.

Applying this to the identity, $F(x_1, \ldots, x_{d-1}, 1 - x_1 - \ldots - x_{d-1}) = 0$ when $x_1, \ldots, x_{d-1} > 0$ and $x_1 + \ldots + x_{d-1} < 1$ means that the identity is still true for all $x_1, \ldots, x_{d-1} \in \mathbb{C}$. Now from the Nullstellensatz on page 164 of [7], we get $(1 - x_1 - \ldots - x_d)$ divides $F(x_1, \ldots, x_d) - 1$ and so

$$F(x_1, \ldots x_d) = g(x_1, \ldots x_d)(1 - x_1 - \ldots - x_d) + 1.$$

1. Clearly,

$$\frac{\partial F}{\partial x_i} = \frac{\partial}{\partial x_i}\left[(1 - x_1 - \ldots - x_d)g(x_1, \ldots, x_d) + 1\right].$$

Now by using the chain rule,

$$\frac{\partial F}{\partial x_i} = -g(x_1, \ldots, x_d) + (1 - x_1 - \ldots - x_d)\frac{\partial g}{\partial x_i}(x_1, \ldots, x_d).$$

When $x_1 + \ldots + x_d = 1$ we have the result

$$\frac{\partial F(x_1, \ldots, x_d)}{\partial x_i} = -g(x_1, \ldots, x_d).$$

Therefore, at points where $x_1 + x_2 + \ldots + x_d = 1$, we get

$$\frac{\partial F(x_1, \ldots, x_d)}{\partial x_i} = \frac{\partial F(x_1, \ldots, x_d)}{\partial x_j} \quad \text{for all } i, j \in \{1, 2, \ldots, d\}. \tag{3.3}$$

2. From the first part of this proof, we have

$$\frac{\partial F}{\partial x_i}(x_1, \ldots, x_d) = -g(x_1, \ldots, x_d) + (1 - x_1 - \ldots - x_d)\frac{\partial g}{\partial x_i}(x_1, \ldots, x_d)$$

and similarly for $\frac{\partial F}{\partial x_j}(x_1, \ldots, x_d)$.

Taking the partial derivatives with respect to l and k such that $k \neq l$, we get

$$\frac{\partial^2 F}{\partial x_i \partial x_k} = -\frac{\partial g}{\partial x_k}(x_1, \ldots, x_d) - \frac{\partial g}{\partial x_i}(x_1, \ldots, x_d) + (1 - x_1 - \ldots - x_d)\frac{\partial^2 g}{\partial x_i \partial x_k}$$

and

$$\frac{\partial^2 F}{\partial x_i \partial x_l} = -\frac{\partial g}{\partial x_l}(x_1, \ldots, x_d) - \frac{\partial g}{\partial x_i}(x_1, \ldots, x_d) + (1 - x_1 - \ldots - x_d)\frac{\partial^2 g}{\partial x_i \partial x_l}.$$

From here, when we set $x_1 + \ldots + x_d = 1$, we have

$$\frac{\partial^2 F}{\partial x_i \partial x_k} = -\frac{\partial g}{\partial x_k}(x_1, \ldots, x_d) - \frac{\partial g}{\partial x_i}(x_1, \ldots, x_d)$$

and

$$\frac{\partial^2 F}{\partial x_i \partial x_l} = -\frac{\partial g}{\partial x_l}(x_1, \ldots, x_d) - \frac{\partial g}{\partial x_i}(x_1, \ldots, x_d).$$

Thus, at the points where $x_1 + \ldots + x_d = 1$, we have

$$\frac{\partial^2 F}{\partial x_i \partial x_k} - \frac{\partial^2 F}{\partial x_i \partial x_l} = -\frac{\partial g}{\partial x_k}(x_1, \ldots, x_d) + \frac{\partial g}{\partial x_l}(x_1, \ldots, x_d).$$

Therefore since the right hand side is independent of x_i, the following formula holds:

$$\frac{\partial^2 F}{\partial x_i \partial x_k} - \frac{\partial^2 F}{\partial x_i \partial x_l} = \frac{\partial^2 F}{\partial x_j \partial x_k} - \frac{\partial^2 F}{\partial x_j \partial x_l}$$

at the points where $x_1 + \ldots + x_d = 1$.

Thus we get

$$\partial_{ik} F + \partial_{jl} F = \partial_{il} F + \partial_{jk} F$$

and if we let $i = l$ and $j = k$, then

$$2\partial_{ij} F = \partial_i^2 F + \partial_j^2 F \tag{3.4}$$

for all i, j at points where $x_1 + x_2 + \ldots + x_d = 1$.

■

Theorem 3.2.6. *If $\mathcal{C}$ is prefix-free and maximal, then*

1. $A_\alpha = A_\beta$

2. $\displaystyle\sum_{w\in\mathcal{C}} \frac{(n_\alpha(w) - n_\beta(w))^2}{d^{l(w)}} = A_\alpha + A_\beta$

for all $\alpha, \beta \in \mathcal{A}$.

Proof. For this theorem we will use Lemma 3.2.4 with F replaced by Ψ.

1. From Lemma 3.2.4, we have $\dfrac{\partial\Psi}{\partial x_i} = \dfrac{\partial\Psi}{\partial x_j}$ when $x_1 + x_2 + \ldots + x_d = 1$. Now calculating the first partial derivative of Ψ with respect to x_i, we get

$$\frac{\partial\Psi}{\partial x_i}(x_1, \ldots, x_d) = \sum_{w\in\mathcal{C}} n_{\alpha_i}(w)\frac{x_1^{\alpha_1(w)} \ldots x_d^{\alpha_d(w)}}{x_i}.$$

Evaluating the first partials of Ψ at $\left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right)$ we now have,

$$\frac{\partial\Psi}{\partial x_i}\left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right) = \sum_{w\in\mathcal{C}} \frac{n_{\alpha_i}(w)}{d^{l(w)-1}} = dA_{\alpha_i}.$$

Therefore, we get that $A_{\alpha_i} = A_{\alpha_j}$.

2. To compute $\dfrac{\partial^2\Psi}{\partial x_i \partial x_j}$, we'll treat the cases $i \neq j$ and $i = j$ separately .

   (a) If $i \neq j$, then

$$\frac{\partial^2\Psi}{\partial x_i \partial x_j} = \sum_{w\in\mathcal{C}} n_{\alpha_i}(w) n_{\alpha_j}(w) \frac{x_1^{n_{\alpha_1}(w)} \ldots x_d^{n_{\alpha_d}(w)}}{x_i x_j}.$$

   Again, evaluating this at $\left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right)$ we get,

$$\frac{\partial^2\Psi}{\partial x_i \partial x_j}\left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right) = \sum_{w\in\mathcal{C}} \frac{n_{\alpha_i}(w) n_{\alpha_j}(w)}{d^{l(w)-2}} = d^2 \sum_{w\in\mathcal{C}} \frac{n_{\alpha_i}(w) n_{\alpha_j}(w)}{d^{l(w)}}. \tag{3.5}$$

(b) If $i = j$

$$\frac{\partial^2 \Psi}{\partial x_i^2} = \sum_{w \in \mathcal{C}} n_{\alpha_i}(w)(n_{\alpha_i}(w) - 1) \frac{x_1^{n_{\alpha_1}(w)} \dots x_d^{n_{\alpha_d}(w)}}{x_i^2}.$$

Again, evaluating this at $\left(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d}\right)$ we get,

$$\frac{\partial^2 \Psi}{\partial x_i^2}\left(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d}\right) = \sum_{w \in \mathcal{C}} \frac{n_{\alpha_i}(w)(n_{\alpha_i}(w) - 1)}{d^{l(w)-2}} = d^2 \sum_{w \in \mathcal{C}} \frac{n_{\alpha_i}(w)(n_{\alpha_i}(w) - 1)}{d^{l(w)}}. \tag{3.6}$$

From Lemma 3.2.4, (3.5) and (3.6), we get

$$\begin{aligned} 2d^2 \sum_{w \in \mathcal{C}} \frac{n_{\alpha_i}(w) n_{\alpha_j}(w)}{d^{l(w)}} &= d^2 \sum_{w \in \mathcal{C}} \frac{n_{\alpha_i}(w)(n_{\alpha_i}(w) - 1)}{d^{l(w)}} + d^2 \sum_{w \in \mathcal{C}} \frac{n_{\alpha_j}(w)(n_{\alpha_i}(w) - 1)}{d^{l(w)}} \\ &= d^2 \left(\sum_{w \in \mathcal{C}} \frac{(n_{\alpha_i}(w))^2}{d^{l(w)}} - \sum_{w \in \mathcal{C}} \frac{n_{\alpha_i}(w)}{d^{l(w)}} + \sum_{w \in \mathcal{C}} \frac{(n_{\alpha_j}(w))^2}{d^{l(w)}} - \sum_{w \in \mathcal{C}} \frac{n_{\alpha_j}(w)}{d^{l(w)}} \right) \end{aligned}$$

Therefore we get

$$\sum_{w \in \mathcal{C}} \frac{n_{\alpha_i}(w)}{d^{l(w)}} + \sum_{w \in \mathcal{C}} \frac{n_{\alpha_j}(w)}{d^{l(w)}} = \sum_{w \in \mathcal{C}} \frac{(n_{\alpha_i}(w))^2}{d^{l(w)}} - 2 \sum_{w \in \mathcal{C}} \frac{n_{\alpha_i}(w) n_{\alpha_j}(w)}{d^{l(w)}} + \sum_{w \in \mathcal{C}} \frac{(n_{\alpha_j}(w))^2}{d^{l(w)}}$$

and thus the following equality is satisfied:

$$\sum_{w \in \mathcal{C}} \frac{(n_{\alpha_i}(w) - n_{\alpha_j}(w))^2}{d^{l(w)}} = A_{\alpha_i} + A_{\alpha_j}. \quad \blacksquare$$

Chapter 4

The equality case of the Kraft-McMillan Inequality

In this section we will look at the equality case of the Kraft-McMillan Inequality. However, we cannot approach this in the same way as the Kraft Inequality.

As the proof for the Kraft-McMillan Inequality uses limits, we will prove the similar results to Kraft's Inequality using an indirect approach.

4.1 Extending our results to uniquely decodable codes

The proposition below is a straightforward generalization of the Kraft-McMillan Inequality.

Proposition 4.1.1. *Let $\phi : \mathcal{A}^* \to [0, \infty)$ be a homomorphism such that*

$$\sum_{\alpha \in \mathcal{A}} \phi(\alpha) = 1. \tag{4.1}$$

Then, if $\mathcal{C}$ is uniquely decodable, we have

$$\sum_{w \in \mathcal{C}} \phi(w) \leq 1. \tag{4.2}$$

Proof. Let $m \in \mathbb{N}$. Then,

$$\begin{aligned}\left(\sum_{w\in\mathcal{C}}\phi(w)\right)^m &= \sum_{w_1,w_2,\ldots,w_m\in\mathcal{C}} \phi(w_1)\phi(w_2)\ldots\phi(w_m) \\ &= \sum_{w_1,w_2,\ldots,w_m\in\mathcal{C}} \phi(w_1 * w_2 * \ldots * w_m) \quad \text{by (2.1)} \\ &\leq \sum_{\substack{v\in\mathcal{A}^* \\ l(v)\leq m\cdot l_{max}}} \phi(v) \quad \text{because } \mathcal{C} \text{ is uniquely decodable} \\ &= \sum_{L=0}^{m\cdot l_{\max}}\left(\sum_{\substack{v\in\mathcal{A}^* \\ l(v)=L}}\phi(v)\right) \\ &= \sum_{L=0}^{m\cdot l_{\max}}\left(\sum_{\alpha\in\mathcal{A}}\phi(\alpha)\right)^L \quad \text{by Lemma (2.2.1)} \\ &= m\cdot l_{\max}+1 \text{ by (4.1) .}\end{aligned} \tag{4.3}$$

Hence $\sum_{m\in\mathcal{C}}\phi(w) \leq \sqrt[m]{m\cdot l_{\max}+1}$, but $\lim_{m\to\infty}\sqrt[m]{m\cdot l_{\max}+1} = 1$. Therefore, $\sum_{m\in\mathcal{C}}\phi(w)\leq 1$ ■

This theorem is Theorem 1 (Kraft Inequality) in [5] with the proof given in the Appendix in loc.cit.

Now we can proceed to prove our main result of this chapter.

Theorem 4.1.2. Assume that $\mathcal{C}$ is uniquely decodable and that $\sum_{w\in\mathcal{C}}\frac{1}{d^{l(w)}} = 1$ (i.e. that equality occurs in the Kraft-McMillan Inequality). For each $\alpha \in \mathcal{A}$,

$$A_\alpha = A_\beta, \quad \text{for all } \alpha,\beta\in\mathcal{A}. \tag{4.4}$$

Proof. Recall the definitions of $\mathcal{H}$,Ψ and ϕ_0 from section 2.2. From Proposition 4.1.1 we know that

$$\Psi(\phi)\leq 1 \quad \text{for all } \phi\in\mathcal{H}. \tag{4.5}$$

Assuming we have equality in the Kraft-McMillan Inequality, we get

$$\Psi(\phi_0) = \sum_{w \in \mathcal{C}} \phi_0(w) = \sum_{w \in \mathcal{C}} \frac{1}{d^{l(w)}} = 1. \tag{4.6}$$

Recall the identification of $\mathcal{H}$ with the $(d-1)$-simplex according to Chapter 2. By (4.5) and (4.6) we get that Ψ attains a maximum at ϕ_0. $\Psi(x_1, x_2, \ldots, x_d)$ has a maximum at $\left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right)$ under the condition $x_1 + x_2 + \ldots + x_d = 1$ and $x_i > 0$ for all $i \in \{1, 2, \ldots, d\}$.

By Remark 2.1, in these coordinates, $\Psi(x_1, x_2, \ldots, x_d) = \sum_{w \in \mathcal{C}} x_1^{n_1(w)} x_2^{n_2(w)} \ldots x_d^{n_d(w)}$, which we've just seen attains a maximum at $(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d})$.

Recall that according to the method of Lagrange Multipliers, if $\phi(x_1, x_2, \ldots, x_d)$ has a local maximum at $(x_1, x_2, \ldots, x_d)$ under the condition $F(x_1, x_2, \ldots, x_d) = 0$, then $\nabla\Psi = \lambda\nabla F$. Since we are looking at points in the neighbourhood of $\left(\frac{1}{d}, \ldots, \frac{1}{d}\right)$ we can ignore $x_i \geq 0$ for the Lagrange multiplier process. In our case, $F(x_1, x_2, \ldots, x_d) = x_1 + x_2 + \ldots + x_d - 1$ and thus we have that

$$\nabla\Psi = \lambda\nabla(x_1 + x_2 + \ldots + x_d - 1) = \lambda(1, 1, \ldots, 1) \text{ at } \left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right).$$

because the gradient of $x_1 + x_2 + \ldots + x_d - 1$ is $(1, 1, \ldots, 1)$. Hence

$$\frac{\partial\Psi}{\partial x_1}\left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right) = \frac{\partial\Psi}{\partial x_2}\left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right) = \ldots = \frac{\partial\Psi}{\partial x_d}\left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right). \tag{4.7}$$

On the other hand, we can compute:

$$\frac{\partial\Psi}{\partial x_1} = \sum_{w \in \mathcal{C}} n_{\alpha_1}(w) x_1^{n_{\alpha_1}(w)-1} x_2^{n_{\alpha_2}(w)} \cdots x_d^{n_{\alpha_d}(w)}$$

$$\frac{\partial\Psi}{\partial x_2} = \sum_{w \in \mathcal{C}} n_{\alpha_2}(w) x_1^{n_{\alpha_1}(w)} x_2^{n_{\alpha_2}(w)-1} \cdots x_d^{n_{\alpha_d}(w)}$$

$$\dots$$

$$\frac{\partial \Psi}{\partial x_d} = \sum_{w \in \mathcal{C}} n_{\alpha_d}(w) x_1^{n_{\alpha_1}(w)} x_2^{n_{\alpha_2}(w)} \cdots x_d^{n_{\alpha_d}(w)-1}$$

Now if we set $x_i = \frac{1}{d}$, we get

$$\begin{aligned} \frac{\partial \Psi}{\partial x_j}\left(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d}\right) &= \sum_{w \in \mathcal{C}} n_{\alpha_j}(w) \frac{1}{d^{l(w)-1}} \\ &= dA_{\alpha_j} \end{aligned} \tag{4.8}$$

By (4.7) and (4.8) it follows that A_{α_j} are all equal. ∎

4.2 Further Directions

We ask how we could take the results of this thesis further.

Question 4.2.1. A prefix-free code $\mathcal{C}$ is maximal if and only if we have equality in the Kraft Inequality. Does this still hold when we broaden our codes from prefix-free to all uniquely decodable codes with equality in the Kraft-McMillan Inequality?

This result is in fact true and a proof of it is given in the Theoretical Computer Science Stack Exchange [6].

Question 4.2.2. Does the refinement on the Kraft Inequality still hold for Kraft-McMillan?

This part will not be pursued in this thesis and is an interesting concept that might be investigated in future work.

Chapter 5

Conclusion

In this paper, we have studied the equality case of two theorems, the Kraft Inequality and McMillan's extension, the Kraft-McMillan Inequality. The goal of the thesis was to study whether prefix-free codes and uniquely decodable codes follow the same rules in terms of character occurrence when restricted to the equality case. Throughout the thesis we went over different extensions to our base theorem stating that when equality occurs in the Kraft Inequality, the weighted number of occurrences of any letter in our alphabet will be equal. We were able to show that this extended into Kraft-McMillan Inequality as well as getting more information on what happens with these weighted number of occurrences for specifically prefix-free codes.

Bibliography

[1] Kraft, Leon G. (1949) *A device for quantizing, grouping, and coding amplitude modulated pulses*, John Wiley & Sons, pp. 108-109.

[2] McMillan, Brockway (1956) *Two inequalities implied by unique decipherability*, IEEE Xplore, IRE Transactions on Information Theory **2** (4) pp.115-116

[3] J. Karush (1961) *A simple proof of an inequality of McMillan*, IRE Trans. Inform. Theory **7** pp. 118

[4] N. Bourbaki (1974) *Elements of Mathematics, Algebra 1*, Herman pp.(12-18)

[5] Hoover H. F. Yin, Ka Hei Ng, Yu Ting Shing, Russel W.F. Lai and Xishi Wang (2019) *Decision Procedure for the Existence of Two-Channel Prefix-Free Codes*, arXiv, https://arxiv.org/abs/1904.12112

[6] (2022) *Theoretical Computer Science Stack Exchange* 51349/maximal-uniquely-decodable-codes

[7] O. Zariski and P. Samuel, Commutative Algebra, Van Nostrand, Princeton (1960), vol II

[8] (2014) *Mathematics Stack Exchange* https://math.stackexchange.com/q/912389

Vita Auctoris

Mr. Xavier Nunes was born in September 1998, Windsor, Ontario, Canada. He graduated from the University of Windsor in 2018 with a Bachelor of Mathematics degree. He decided to pursue in education at the same university and stay in his hometown. He is currently a candidate for the Master of Science degree in Mathematics at the University of Windsor