trim

TEXTS AND READINGS IN MATHEMATICS 45

Coding Theorems of Classical and Quantum Information Theory

Second Edition

K. R. Parthasarathy

HINDUSTAN BOOK AGENCY

Coding Theorems of Classical and Quantum Information Theory

Second Edition

K. R. Parthasarathy
Indian Statistical Institute
New Delhi

HINDUSTAN BOOK AGENCY

Published by

Hindustan Book Agency (India)
P 19 Green Park Extension
New Delhi 110 016
India

email: info@hindbook.com
www.hindbook.com

ISBN 978-93-80250-41-0

TEXTS AND READINGS IN MATHEMATICS 45

Coding Theorems of Classical and Quantum Information Theory

Second Edition

Coding Theorems of Classical and Quantum Information Theory

Second Edition

K. R. Parthasarathy
Indian Statistical Institute
New Delhi

HINDUSTAN BOOK AGENCY

Published by

Hindustan Book Agency (India)
P 19 Green Park Extension
New Delhi 110 016
India

email: info@hindbook.com
www.hindbook.com

ISBN 978-93-80250-41-0

To

Shyama

Preface

The logarithmic connection between entropy and probability was first enunciated by L.E. Boltzmann (1844-1906) in his kinetic theory of gases. His famous formula for entropy S is $S = k \log W$ (as engraved on his tombstone in Vienna) where k is a constant and W is the number of possible microstates corresponding to the macroscopic state of a system of particles in a gas. Ignoring the constant k and replacing $\log W$ by $-\log P(E)$ where $P(E)$ is the probability of an event E in the probability space $(\Omega, \mathcal{F}, P)$ of a statistical experiment, C. E. Shannon (1916–2001) looked upon $-\log P(E)$ as a measure of the information gained about the probability space from the occurrence of E. If X is a simple random variable on this probability space assuming the values $a_1, a_2, ..., a_k$ from a finite set with $P(X = a_j) = p_j$ for each j then the famous Shannon entropy $H(X) = -\sum_j p_j \log\ p_j$ is the expected information about $(\Omega, \mathcal{F}, P)$ gained from observing X. Centred around this idea of entropy a mathematical theory of communication was woven by Shannon in a celebrated pair of papers in the 1948 volume of the Bell System Technical Journal. Here Shannon established two fundamental coding theorems about the optimal compressibility of a text in its storage and the optimal capacity of a channel in communicating a text after encoding.

The modern approach to information theory is to view a text in any alphabetic language as a finite time realization of a stochastic process in discrete time with values in a finite set (called alphabet) and consider the quantity $-\frac{1}{n} \log P(x_0, x_1, ..., x_{n-1})$ as the rate at which information is generated by the text $x_0, x_1, ..., x_{n-1}$ during the period $[0, n-1]$. Under fairly general conditions this rate exhibits an asymptotic stability property as n becomes large. Through the papers of B. Mcmillan, A. Feinstein, L. Breiman, J. Wolfowitz and others it is now known that an appeal to this stability property enlarges the scope of Shannon's coding theorems. This gets enriched further by exploiting the Kryloff-Bogoliouboff theory of disintegrating an invariant probability measure into its ergodic components. The first three chapters of this little book are devoted to Shannon's coding theorems and their enriched versions. However, we have not touched upon the coding theorems in their most general form as presented in the book of Te Sun Han [14].

A decade after the appearance of Shannon's famous work, A. N. Kolmogorov (1903–1987) demonstrated, rather dramatically, how the notion of the expected rate of generation of entropy or information assumes an intelligence of its own and yields a nonspectral invariant for the classification of dynamical systems. Since very little extra effort is involved in presenting this beautiful work I have taken the liberty of including it as a small digression.

In 1932, while laying the mathematical foundations for quantum mechanics, John von Neumann (1903–1957) introduced the fruitful notion of entropy for the state of a quantum system. If ρ is the density operator of the state of a quantum system then its von Neumann entropy $S(\rho)$ is defined by $S(\rho) = -\mathrm{Tr}\,\rho \log \rho$. Through the work of A. S. Holevo, B. Schumacher, W. D. Westmoreland and others as outlined in the book of Nielsen and Chuang [24] the reader can recognize the role of von Neumann entropy in attempts to formulate and establish quantum versions of the coding theorems of Shannon when classical messages are encoded as quantum states and decoding is done by generalized measurements. Our last and the fourth chapter is devoted to a self-contained account of these coding theorems in the quantum avatar as described in the elegant work of A.Winter in his 1999 paper [48].

A large part of the first three chapters of this book does not use anything more than Chebyshev's inequality. The ergodic theorem, martingale theorem and decomposition of an invariant probability measure into its ergodic components are used in arriving at the more sophisticated versions of the classical coding theorems. The last chapter demands nothing more than a knowledge of operators in a finite dimensional Hilbert space.

The present exposition has evolved through the courses of lectures I had given at the Indian Statistical Institute, Calcutta in 1961, the Tata Institute of Fundamental Research, Mumbai in 2001 and 2002, the Institute of Mathematical Sciences, Chennai in 2001 and 2005, the Ramanujan Institute of Advanced Study in Mathematics at the University of Madras in 2005 and Chungbuk National University, Cheongju, Korea in 2005. I am grateful to C. R. Rao who suggested to me in 1959 the study of information theory for my PhD thesis and J. Radhakrishnan, R. Parimala, R. Balasubramanian, M. Krishna, V. Arvind, S. Parvathi, K. Parthasarathy, V. Thangaraj and Un Cig Ji who were instrumental in organising these lectures in a congenial atmosphere. I thank Anil Shukla for his elegant TEX of my notes with patience in spite of my repeated requests for changes and corrections. Thanks to the careful proof-reading by P. Vanchinathan a significant control over the number of grammatical, typographical and TEX errors has been exercised. The support given by my colleagues at the Delhi Centre of the Indian Statistical Institute is gratefully acknowledged.

Indian Statistical Institute
Delhi Centre
New Delhi - 110 016
India

K. R. Parthasarathy
January 2007

Preface to the revised edition

The essential feature of the revised edition is the inclusion of a new chapter devoted to the Knill-Laflamme theory of quantum error correction and its consequences in the construction of t-error correcting quantum codes. Our approach is based on the unification of classical and quantum error correcting codes through imprimitivity systems for finite group actions.

Many typographical error corrections and some minor changes have been made in the text of the first edition.

I have greatly benefited from discussions with V. Arvind and Harish Parthasarathy. Ajit Iqbal Singh has rendered valuable help in carefully reading the manuscript and suggesting many improvements. Anil Kumar Shukla has Texed the revised manuscript showing tremendous patience in fulfilling my requests for repeated changes in the text. The continued support of my colleagues in the institute has enabled the completion of this revision in reasonable time. To all of them I express my sincere thanks.

Indian Statistical Institute
Delhi Centre
New Delhi - 110 016
India

K. R. Parthasarathy
September 2012

Contents

Chapter 1

Entropy of Elementary Information Sources

1.1 Uniquely decipherable and irreducible codes

We begin with an elementary analysis of maps from a finite set into the free semigroup generated by another finite set and develop a terminology appropriate to information theory.

Consider any finite set A of cardinality a denoted as $\#A = a$. We say that A is an *alphabet* of *size* a and call any element x in A as a *letter* from the alphabet A. Any element $w = (x_1, x_2, \ldots, x_n)$ in the n-fold cartesian product A^n of copies of A is called a *word* of *length* n, the latter denoted by $l(w)$. It is customary to express such a word as $w = x_1 x_2 \ldots x_n$ by dropping the brackets and commas. Denote

$$S(A) = \bigcup_{r=1}^{\infty} A^r$$

and for any $w_1 = x_1 x_2 \ldots x_{n_1} \in A^{n_1}$, $w_2 = y_1 y_2 \ldots y_{n_2} \in A^{n_2}$ define the *product word* $w_1 w_2$ by $w_1 w_2 = x_1 x_2 \ldots x_{n_1} y_1 y_2 \ldots y_{n_2}$. Thus $l(w_1 w_2) = l(w_1) + l(w_2)$. Clearly, this multiplication is associative. It makes $S(A)$ a semigroup without an identity element. We call $S(A)$ the *free semigroup* or *word semigroup* generated by the alphabet A.

Let A, B be alphabets of sizes a, b respectively. A one-to-one (or injective) map $f : A \to S(B)$ is called a *code* with *message alphabet* A and *encoding alphabet* B. When B is the two point set $\{0, 1\}$ such a code f is called a *binary code.* Any word in the range of a code f is called a *basic code word.* Start with a code $f : A \to S(B)$ and extend it uniquely to a map $\tilde{f} : S(A) \to S(B)$ by putting

$$\tilde{f}(w) = \tilde{f}(x_1 x_2 \ldots x_n) = f(x_1) f(x_2) \ldots f(x_n)$$

for any word $w = x_1 x_2 \ldots x_n$ in $S(A)$. Then f is said to be a *uniquely decipherable code* if its extension $\tilde{f}$ is also one to one. The code f is said to be *irreducible* if for any two letters x and y in A, $f(y) \neq f(x)$ and $f(y)$ cannot be expressed as $f(y) = f(x)w$ for any word w in $S(B)$. A simple examination shows that an irreducible code is uniquely decipherable.

We shall now establish a necessary condition for a code $f : A \to S(B)$ to be uniquely decipherable.

Theorem 1.1.1 (Sardinas and Patterson [40]) Let A, B be alphabets of sizes a, b respectively and let $f : A \to S(B)$ be a uniquely decipherable code. Then

$$\sum_{x \in A} b^{-l(f(x))} \leq 1. \tag{1.1.1}$$

where $l(w)$ denotes the length of the word w.

Proof. Let

$$\begin{aligned} L &= \max \{ l(f(x)) \mid x \in A \}, \\ c_r &= \# \{ x \mid l(f(x)) = r \}. \end{aligned}$$

Then the left hand side of (1.1.1) can be expressed as

$$\begin{aligned} \sum_{x \in A} b^{-l(f(x))} &= \sum_{r=1}^{L} \sum_{x : l(f(x)) = r} b^{-l(f(x))} \\ &= \sum_{r=1}^{L} c_r b^{-r} \\ &= P(b^{-1}) \end{aligned}$$

where P is the polynomial defined by

$$P(z) = \sum_{r=1}^{L} c_r z^r.$$

Define

$$N(k) = \# \left\{ \tilde{f}(w) \mid w \in S(A), \quad l(\tilde{f}(w)) = k \right\},$$

the cardinality of the set of all code words of length k. Clearly, $N(k) \leq b^k$ for $k = 1, 2, \ldots$ Thus the power series

$$F(z) = 1 + \sum_{k=1}^{\infty} N(k) z^k$$

converges to an analytic function in the open disc $\left\{ z \mid \ |z| < b^{-1} \right\}$. Introduce the convention that $N(0) = 1$ and $N(k) = 0$ if $k < 0$. Since every code word