Information, Codes and Ciphers

By Jeremy Le for MATH3411 24T3

1 Introduction

1.1 Mathematical Model

To give a mathematical framework for digital data transmission, define

- a **source alphabet** $S = \{s_1, s_2, \ldots, s_q\}$ of q symbols
- a **code alphabet** A of r symbols probabilities $p_i = P(s_i)$
- a **code** that encodes each symbol s_i by a codeword which is a **string** of code symbols.

1.2 Assumed Knowledge

- Modular Arithmetic and the Division Algorithm
- Probability (Binomial Distribution and Bayes' Rule)
- Linear Algebra (Linear combination, independence, etc...)

1.3 Morse Code

Morse code is a **ternary** code (radix 3). Its alphabet is

1. • called **dot**
2. — called **dash**
3. `p` a **pause**

The codewords are strings of • and — **terminated** by `p`.

1.4 ASCII

American National **S**tandard **C**ode for **I**nformation **I**nterchange.

Binary code of fixed codeword length, namely 7, with $2^7 = 128$ encoded symbols.

The extended ASCII is a code like the 7-bit ASCII but with an extra bit in the front used as a check bit, requiring the number of 1's to be even.

1.5 ISBN

International **S**tandard **B**ook **N**umber.

They have 10 bits, with it's last bit being a check bit, requiring

$$\sum_{i=1}^{10} ix_i \equiv 0 \pmod{11}.$$

2 Error Detection and Correction Codes

2.1 ISBN-10 Error Capability

ISBN-10 numbers are capable of detecting the two types of errors:

1. getting a digit wrong
2. interchanging two (unequal) digits.

2.2 Binary Repetition Codes

The binary $(2t+1)$-repetition code is t-error correcting.
The binary $2t$-repetition code is $(t-1)$-error correcting and t-error detecting.