MATH 525
Sections 1.1–1.6 – Basics of Coding Theory

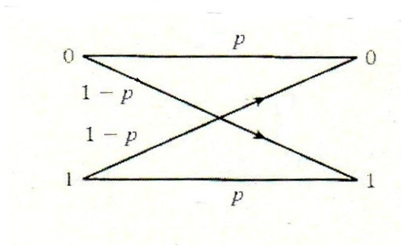
August 26, 2020

Definitions and Assumptions

- Digits or Bits: $0, 1$.
- Word: Sequence of digits, e.g., $001, 100110, 1101$, etc.
- Length of a word: # of digits it has.
- Channel: The physical link that connects a data source to a data sink. In this course, it will represent the theoretical channel model with certain error characteristics.
- Binary channel: Only 0s and 1s and are transmitted or received over it.
- Binary code: Set of words. Example: $\{00, 110, 01, 11\}$ is a code.
- Block code: All words in it have the same length, called the length of the code. Example: $\{000, 111\}$ is a code of length 3.
- Repetition and parity-check codes: They are both block codes. More details during the lecture.
- The words that belong to a code are called codewords. The number of words in a code C is denoted by $|C|$. That number is known as the size of the code or the cardinality of the code.
- We will assume, initially, that no digits become lost. If a word of length n is transmitted, then a corresponding word of the same length is received.

Definitions and Assumptions - Cont'd.

- We will assume that errors occur independently, that is, the occurrence of error during a time slot does not imply anything about what will happen during the next time slot. An important situation where this is not true is when errors occur in *bursts* (Chapter 7).
- The **Binary Symmetric Channel** (BSC):



  - The error probability is the same, regardless of whether 0 or 1 is transmitted; p is known as the reliability of the channel.
  - Special cases: $p = 1$ and $p = 0$.
  - We can always assume that $\frac{1}{2} \leq p < 1$.

Definitions and Assumptions - Cont'd.

Definition

The information rate of a code C is the proportion of digits that convey information. Formally, it is defined as

$$R = \frac{\log_2 |C|}{n} \text{ bits per block}$$

where n is the length of C.

For example, $C = \{000, 111\}$ has a rate of $\frac{1}{3}$.

Error Detection and Error Correction

- When a received word is not a codeword, we say that the code has detected that errors occurred during the transmission.
- Correcting errors means converting a received word (in error) into a codeword. Usually, the received word is converted into the most likely codeword transmitted.

Example

$C_1 = \{00, 01, 10, 11\}$ cannot detect any errors, let alone correct any errors.

Example

$C_2 = \{000, 011, 101, 110\}$ (parity-check code of length 3) can detect one error (affecting any codeword).

Example

A repetition code. $C_3 = \{000000, 010101, 101010, 111111\}$ can detect up to two errors (affecting any codeword). Suppose 110101 is received. The most likely codeword transmitted is 010101. So we correct 110101 to 010101.

Finding the most likely codeword transmitted

Problem: Given a received word w, how can we decide between v_1 and v_2 (as sent codewords)?

Let $\phi_p(v, w) =$ probability of receiving w given that v was sent. We have:

$$\phi_p(v, w) = p^{n-d} q^d$$

where $q = 1 - p$ and d is the number of positions in which v and w disagree.

Theorem

Suppose we have a BSC with $\frac{1}{2} \leq p < 1$. Suppose v_1 and w disagree in d_1 positions and v_2 and w disagree in d_2 positions. Then

$$\phi_p(v_1, w) \leq \phi_p(v_2, w) \iff d_1 \geq d_2.$$

Conclusion: Given a received word r, the decoder will look for the codeword that least disagrees with r and will decode r into it.