Coding and Information Theory
Chapter 5
Using an Unreliable Channel - B

Xuejun Liang

Fall 2022

Chapter 5
Using an Unreliable Channel

1. Decision Rules
2. An Example of Improved Reliability
3. Hamming Distance
4. Statement and Outline Proof of Shannon's Theorem
5. The Converse of Shannon's Theorem
6. Comments on Shannon's Theorem

Quick Review of Last Lecture

- Decision Rules: $b_j \rightarrow \Delta(b_j) = a_{j^*}$

$$\mathrm{Pr_C} = \sum_j q_j Q_{j^*j} = \sum_j R_{j^*j}$$

$$\mathrm{Pr_E} = 1 - \mathrm{Pr_C} = 1 - \sum_j R_{j^*j} = \sum_j \sum_{i \neq j^*} R_{ij}$$

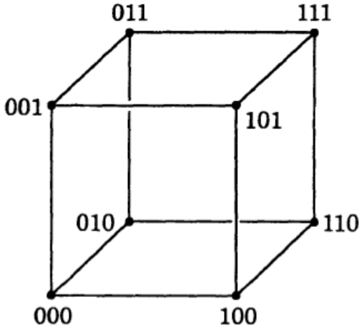
  - Ideal observer rule using $R_{i,j}$ or $Q_{i,j}$
  - Maximum likelihood rule using $P_{i,j}$.
  - Examples
- An Example of Improved Reliability
  - Sending each input symbol a = 0 or 1 three times in succession

    $\mathrm{Pr}_E = 3PQ^2 + Q^3 = Q^2(3 - 2Q) \approx 3Q^2$
  - Any subset $C \subseteq A^n$ can be used for a set of code-words.
  - The transmission rate

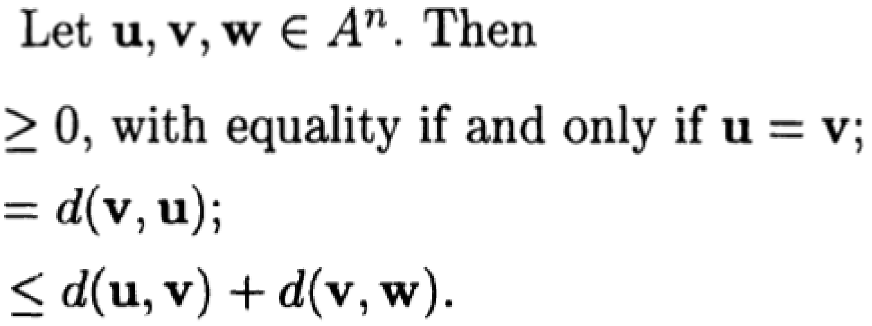
$$R = \frac{\log_r |\mathcal{C}|}{n}$$

5.3 Hamming Distance

- Let $\boldsymbol{u} = u_1 \ldots u_n$ and $\boldsymbol{v} = v_1 \ldots v_n$ be words of length n in some alphabet A, so $\boldsymbol{u}, \boldsymbol{v} \in A^n$. The Hamming distance d($\boldsymbol{u}, \boldsymbol{v}$) between $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined to be the number of subscripts i such that $u_i \neq v_i$.
  - $d(u, v) = |\{i \mid u_i \neq v_i\}|$
- Example 5.6
  - Let $\boldsymbol{u}$ = 01101 and $\boldsymbol{v}$ = 01000 in Z_2^5. Then d($\boldsymbol{u}, \boldsymbol{v}$) = 2.
- Example 5.7
  - We can regard the words in Z_2^3 as the eight vertices of a cube.



Hamming Distance (Cont.)

- Lemma 5.8 

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in A^n$. Then

(a) $d(\mathbf{u}, \mathbf{v}) \geq 0$, with equality if and only if $\mathbf{u} = \mathbf{v}$;

(b) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$;

(c) $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$.

  - Proof of (c)

$$\{i \mid u_i \neq w_i\} \subseteq \{i \mid u_i \neq v_i\} \cup \{i \mid v_i \neq w_i\}$$

$$|\{i \mid u_i \neq w_i\}| \leq |\{i \mid u_i \neq v_i\}| + |\{i \mid v_i \neq w_i\}|$$

- To transmit information through Γ, we choose a code $C \subseteq A^n$ for some n, and use the maximum likelihood decision rule.
  - Decode each received word as the code-word most likely to have caused it. (Using forward probability P_{ij}.)

Hamming Distance (Cont.)

- For simplicity, assume that Γ is the BSC, with P > $1/2$, so A = B = Z_2 and r = 2.
  - The **maximum likelihood** decision rule means for any output $v \in Z_2^n$, we decode v as the code-word u = Δ(v) ∈ C which maximizes the forward probability Pr(v I u).
  - Note: a code-word u which maximizes Pr(v I u) is one which minimizes d(u, v).

    If $d(u, v) = i$ then

    $$\Pr(v|u) = Q^i P^{n-i} = P^n \left(\frac{Q}{P}\right)^i$$

  - So, this is also called the **nearest neighbor decoding**

5.4 Statement of Shannon's Theorem

- Informally
  - Shannon's Theorem says that if we use long enough code-words then we can send information through a channel Γ as accurately as we require, at a rate arbitrarily close to the capacity C of Γ.
- Theorem 5.9
  - Let Γ be a binary symmetric channel with P > $1/2$, so Γ has capacity C = 1 - H(P) > 0, and let $\delta, \varepsilon > 0$. Then for all sufficiently large n there is a code $C \subseteq Z_2^n$, of rate R satisfying $\mathrm{C} - \varepsilon \leq R < \mathrm{C}$, such that nearest neighbor decoding gives error-probability $\mathrm{Pr}_E < \delta$.

Outline Proof of Shannon's Theorem

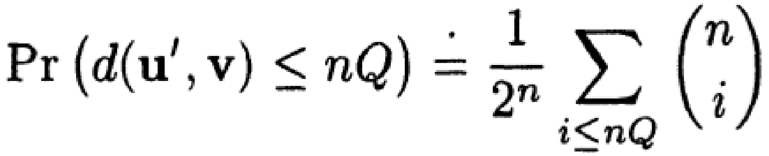
- Let R < C, Randomly chose $C \subset Z_2^n$, $|C| = 2^{nR}$.
- Rate of C = ${}^{log_2 2^{nR}}/_{n} = R$
- Sending $\boldsymbol{u}$, expect to receive $\boldsymbol{v}$ such that d($\boldsymbol{u}$, $\boldsymbol{v}$) $\approx nQ$
- Receiving $\boldsymbol{v}$, decode $\Delta(\boldsymbol{v}) = \boldsymbol{u}$ such that d($\boldsymbol{u}$, $\boldsymbol{v}$) $\approx nQ$
- Using the nearest neighbor rule, if decoding is incorrect then there must be some $\boldsymbol{u}$' $\neq \boldsymbol{u}$ in C with d($\boldsymbol{u}$',$\boldsymbol{v}$) $\leq$ d($\boldsymbol{u}$,$\boldsymbol{v}$).
- So

$$\mathrm{Pr_E} \leq \sum_{\mathbf{u}' \neq \mathbf{u}} \Pr\left(d(\mathbf{u}', \mathbf{v}) \leq nQ\right), \qquad (5.4)$$

- The upper bound on Pr_E in (5.4) is equal to

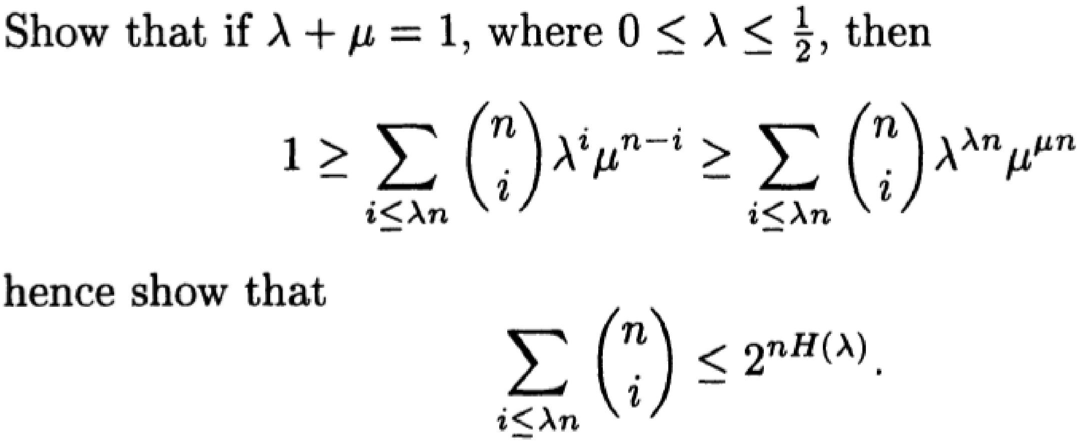
$$(|\mathcal{C}| - 1) \Pr\left(d(\mathbf{u}', \mathbf{v}) \leq nQ\right) < 2^{nR} \Pr\left(d(\mathbf{u}', \mathbf{v}) \leq nQ\right).$$

- For any given $\boldsymbol{v}$ and i, $|\{\, \boldsymbol{u}' \in Z_2^n : \mathrm{d}(\boldsymbol{u}', \boldsymbol{v}) = i \,\}| = \binom{n}{i}$
- So, $|\{\boldsymbol{u}' \in Z_2^n : \mathrm{d}(\boldsymbol{u}', \boldsymbol{v}) \leq nQ\}| = \sum_{i \leq nQ} \binom{n}{i}$
- Therefore



$$\Pr\left(d(\mathbf{u}', \mathbf{v}) \leq nQ\right) \doteq \frac{1}{2^n} \sum_{i \leq nQ} \binom{n}{i}$$

- Exercise 5.7



Show that if $\lambda + \mu = 1$, where $0 \leq \lambda \leq \frac{1}{2}$, then

$$1 \geq \sum_{i \leq \lambda n} \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i \leq \lambda n} \binom{n}{i} \lambda^{\lambda n} \mu^{\mu n}$$

hence show that

$$\sum_{i \leq \lambda n} \binom{n}{i} \leq 2^{nH(\lambda)}.$$

Outline Proof (Cont.)

- Putting $\lambda = Q$ in Exercise 5.7, we have

$$\sum_{i \le nQ} \binom{n}{i} \le 2^{nH(Q)}$$

- Thus (5.4) becomes

$$\mathrm{Pr_E} < 2^{nR} . \frac{1}{2^n} . 2^{nH(Q)} = 2^{n(R-1+H(Q))} = 2^{n(R-C)}$$

  - Note: C = 1 - H(P) = 1 - H(Q).

- Now R < C, so $2^{n(R-C)} \to 0$ as $n \to \infty$, and hence $\mathrm{Pr}_E \to 0$ also.

5.5 The Converse of Shannon's Theorem

- Informally
  - The converse of Shannon’s Theorem says that one can not do better than what the Shannon’s Theorem says.
- The converse of Shannon’s Theorem
  - If C' > C then it is not true that for every $\varepsilon > 0$ there is a sequence of codes C, of lengths $n \to \infty$, and of rates R satisfying C' - $\varepsilon \leq$ R < C', such that $\Pr_E \to 0$ as $n \to \infty$.
- The Fano bound
  - gives a lower bound on the error-probability. (See Theorem 5.10 on the next slide.)

The Fano Bound

- Theorem 5.10
  - Let Γ be a channel with input A and output B. Then the error-probability Pr_E corresponding to any decision rule Δ for Γ satisfies

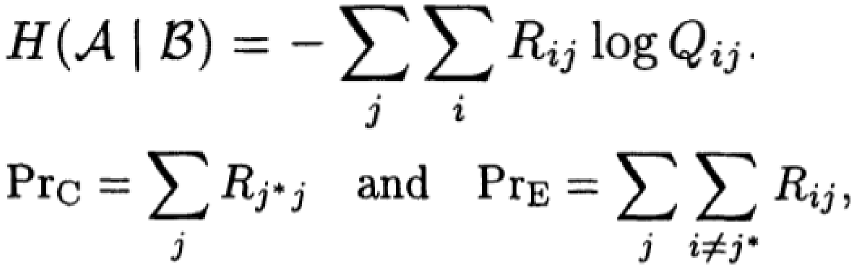
$$H(\mathcal{A} \mid \mathcal{B}) \leq H(\mathrm{Pr_E}) + \mathrm{Pr_E} \log(r-1) \qquad (5.5)$$

  where r is the number of symbols in A

- Meaning of inequality (5.5)
  - Given b_j, the receiver decodes $a_{j*} = \Delta(b_j)$, which may or may not be the actual symbol a_i transmitted.
  - The left-hand side of (5.5) is the extra information the receiver needs (on average) in order to know a_i

The Fano Bound (Cont.)

- Meaning of inequality (5.5)
  - This extra information can be divided into two parts:
    a) Whether or not decoding is correct, that is, whether or not $a_{j*} = a_i$;
    b) If decoding is incorrect, then which $a_i (i \neq j^*)$ out of r-1 symbols was transmitted.
  - The information in (a) has value $H(\mathrm{Pr}_E)$
  - The information in (b) has value at most $\mathrm{Pr}_E \log(r-1)$
- Note: we have



$$H(\mathcal{A} \mid \mathcal{B}) = -\sum_j \sum_i R_{ij} \log Q_{ij} .$$

$$\mathrm{Pr}_{\mathrm{C}} = \sum_j R_{j^* j} \quad \text{and} \quad \mathrm{Pr}_{\mathrm{E}} = \sum_j \sum_{i \neq j^*} R_{ij} ,$$

Examples

- Example 5.11
  - Let Γ be the BSC , and as a rather extreme example of a code let us take C = A^n, so R = 1.
  - If 0 < P < 1 we have C = 1 - H(P) < 1, so R > C.
  - Using the identity function $\Delta(u)$ = u as a decision rule, we see that decoding is correct if and only if there are no errors, so $\mathrm{Pr}_E = 1 - P^n \rightarrow 1$ as $n \rightarrow \infty$.

Examples (Cont.)

- Example 5.12
  - The Hamming codes of length n of the form $2^c - 1$ and rate R = $(n$ - $c)/n$, so R → 1 as n → ∞.
  - If we use a BSC with 0 < P < 1, then C = 1 - H(P) < 1 and hence R > C for all sufficiently large n.
  - The nearest neighbor decoding is correct if and only if there is at most one error (shall see this in §7.4), so $\mathrm{Pr}_E = 1 - P^n - nP^{n-1}Q \to 1$ as $n \to \infty$.

5.6 Comments on Shannon's Theorem

- Theorem 5.13 (The general form of Shannon's Theorem)
  - Let Γ be an information channel with capacity C > 0, and let $\delta, \varepsilon > 0$. For all sufficiently large n there is a code C of length n, of rate R satisfying $\mathrm{C} - \varepsilon \leq R < \mathrm{C}$, together with a decision rule which has error-probability $\mathrm{Pr}_E < \delta$.
- Comment 5.14
  - In order to achieve values of R close to C and Pr_E close to 0, one may have to use a very large value of n.
  - This means that code-words are very long, so encoding and decoding may become difficult and time-consuming.

Comments on Shannon's Theorem

- Comment 5.14
  - Moreover, if n is large then the receiver experiences delays while waiting for complete codewords to come through; when a received word is decoded, there is a sudden burst of information, which may be difficult to handle.
- Comment 5.15
  - Shannon's Theorem tells us that good codes exist, but neither the statement nor the proof give one much help in finding them.

Comment 5.15 (Cont.)

- The proof shows that the "average" code is good, but there is no guarantee that any specific code is good: this has to be proved by examining that code in detail.
- One might choose a code at random, as in the proof of the Theorem, and there is a reasonable chance that it will be good.
- However, random codes are very difficult to use: ideally, one wants a code to have plenty of structure, which can then be used to design effective algorithms for encoding and decoding.
- We will see examples of this in Chapters 6 and 7, when we construct specific codes with good transmission rates or error-probabilities.