Lecture 10 — Feb 9

Lecturer: David Tse *Scribe: Daria R, Shiv K, Yuki N, Nipun A, Vivek B*

10.1 Outline

- Fano's Inequality
- Jointly typical sequences
- Getting to capacity: Sphere packing with random spheres

10.1.1 Readings

- CT: 7.6, 7.7, 13.7

10.2 Recap - Converse to the Coding Theorem

Figure 10.1: Channel Coding Model.

For the channel coding model (Figure 10.1), we proved

$$H(W|\hat{W}) \geq n(R-C). \tag{10.1}$$

Therefore, for $R > C$, the original message W, conditioned on $\hat{W}$ has large uncertainty; which intuitively suggests that p_e must also be large. Now, we rigorously prove this claim using *Fano's Inequality.*

Theorem 1. (*Fano's Inequality*) For an **estimator** $\hat{U}$, and random variables U, V such that $U \to V \to \hat{U}$,

$$P(\hat{U} \neq U) \frac{\geq H(U|\hat{U}) - 1}{\log |\mathcal{U}|}.$$

Proof. Refer CT. □

Applying Fano's inequality on Markov chain $W \to X^n \to Y^n \to \hat{W}$ (refer Figure 10.1), we obtain

$$\begin{aligned} P(\hat{W} \neq W) &\geq \frac{H(W|\hat{W}) - 1}{\log W} \\ &\geq \frac{n(R-C) - 1}{nR} \\ &= 1 - \frac{C}{R} - \frac{1}{nR}. \end{aligned}$$

This equation shows that if $R > C$, the probability of error is bounded away from 0 for sufficiently large n. Hence, we cannot achieve arbitrary low probability of error for $R > C$.

10.2.1 Data processing

In the last lecture, we used data processing theorem to derive the equation (10.1).

Theorem 2. Suppose random variables X, Y, and Z form a Markov Chain (i.e. $X \to Y \to Z$), then

$$I(X;Y) \geq I(X;Z).$$

Proof. Using the chain rule for mutual information, expand $I(X;Y,Z)$ in two following ways:

$$\begin{aligned} I(X;Y,Z) &= I(X;Y) + I(X;Z|Y) \\ I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \end{aligned}$$

Since X, Y, and Z form a Markov chain, $I(X;Z|Y) = 0$. From the non-negativity of mutual information we know that $I(X;Y|Z) \geq 0$. Thus

$$\begin{aligned} I(X;Y) &= I(X;Y,Z) \\ &= I(X;Z) + I(X;Y|Z) \\ &\geq I(X;Z) \end{aligned}$$

□

10.3 Sphere Packing

Our discussions in this section will hold true for a general channel but we will use $BSC(p)$ as a running example for concreteness.

A code $\mathcal{C}$ maps every message in $\{0,1\}^{nR}$ to a **codeword** in the space $\{0,1\}^n$. Since, $BSC(p)$ introduces roughly np errors, the natural optimization problem is - **(i)** Maximize the number of codewords, **(ii)** subject to the **constraint** that every pair of codewords is at least $2np$ apart in Hamming distance.

This problem is known as the 'Sphere packing problem' (refer Figure **??**) and till date it remains unsolved!

Figure 10.3

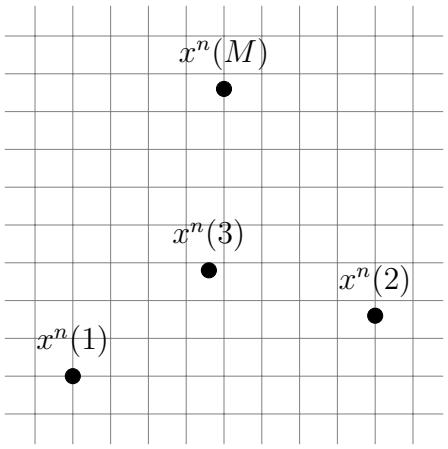

Figure 10.2: Sphere Packing problem : How to pack codewords $x^n(1),\ x^n(2), \cdots\ x^n(M)$, such that hamming distance between every pair of codewords is at least $2np$?

Sphere packing is a very hard combinatorial problem because of its strict constraints. We instead solve a relaxed version where we pack spheres of radius np which satisfy the above constraint with high probability.

10.3.1 Jointly Typical sequences

For ease of notation, let us denote the sequence $(x_1, x_2, \cdots, x_n)$ by x^n, i.e. $x^n \triangleq (x_1, x_2, \cdots, x_n)$.

Definition 1. For i.i.d. random variables (X^n, Y^n), (x^n, y^n) is a **jointly typical** sequence, if it satisfies

$$\begin{aligned} p(x^n) &\sim 2^{-nH(X)} \\ p(y^n) &\sim 2^{-nH(Y)} \\ p(x^n, y^n) &\sim 2^{-nH(X,Y)}. \end{aligned}$$

Example 1. For a random variable $X \sim Bern(1/2)$ and $BSC(p)$, we have

$$H(X, Y) = H(Y|X) + H(X) = H(p) + 1.$$

Let (x^n, y^n) be jointly typical, then

$$\begin{aligned} p(x^n) &\sim 2^{-n} \\ p(y^n) &\sim 2^{-n} \\ p(x^n, y^n) &\sim 2^{-n(H(p)+1)}, \end{aligned}$$

and also

- $\left|\left\{x^n : p(x^n) = 2^{-n}\right\}\right| : \sim 2^n$.
- $\left|\left\{y^n : p(y^n) = 2^{-n}\right\}\right| : \sim 2^n$.
- $\left|\left\{x^n, y^n : p(x^n, y^n) = 2^{-n(H(p)+1)}\right\}\right| : \sim 2^{n(H(p)+1)}$ $\Big(< 2^{2n} \Big)$.

Conditional distribution

For jointly typical sequences (x^n, y^n),

$$\begin{aligned} p(y^n|x^n) &= \frac{p(x^n, y^n)}{p(x^n)} \\ &\sim \frac{2^{-nH(X,Y)}}{2^{-H(X)}} \\ &= 2^{-n\left(H(X,Y)-H(X)\right)} \\ p(y^n|x^n) &\sim 2^{-nH(Y|X)}. \end{aligned} \tag{10.2}$$

In Example **??**, $p(y^n|x^n) = 2^{-nH(p)}$.

10.3.2 Sphere packing with random spheres

From equation (10.2), we see that the "noise sphere" around each codeword x^n, contains around $2^{nH(Y|X)}$ typical sequences y^n. Since there are $2^{H(Y)}$ typical sequences y^n in the whole space $\{0,1\}^n$, the number of disjoint noise spheres (each corresponding a single codeword) is upper bounded by

$$\frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{nI(X;Y)}.$$

Therefore, the total number of codewords is also upper bounded,

$$M = 2^{nR} \leq 2^{nI(X;Y)}.$$

Taking the log of both sides of the above inequality and dividing by n leads to

$$R \leq I(X;Y) \leq C.$$