LECTURE 10

Last time:

- Maximizing capacity: Arimoto-Blahut
- Examples

Lecture outline

- The channel coding theorem overview
- Upper bound on the error probability
- Bound on not being typical
- Bound on too many elements being typical
- Coding theorem (weak)

Reading: Reading: Scts. 8.4, 8.7.

Overview

Consider a DMC with transition probabilities $P_{Y|X}(y|x)$

For any block length n, let

$$P_{\underline{Y}^n|\underline{X}^n}(\underline{y}^n|\underline{x}^n) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$$

$$P_{\underline{X}^n}(\underline{x}^n) = \prod_{i=1}^{n} P_X(x_i)$$

$$P_{\underline{Y}^n}(\underline{y}^n) = \prod_{i=1}^{n} P_Y(y_i)$$

Let R be an arbitrary rate $R < C$

For each n consider choosing a code of $M = \lfloor e^{nR} \rfloor$ codewords, where each codeword is chosen independently with probability assignment $P_{\underline{X}^n}(\underline{x}^n)$

We assume the messages are equiprobable, so the entropy rate (per symbol) of the messages is R

Overview

Let $\epsilon = \frac{C-R}{2}$ and let the set T_ϵ^n be the set of pairs $(\underline{x}^n, \underline{y}^n)$ such that

$$|\frac{1}{n} i\left(\underline{x}^n; \underline{y}^n\right) - C| \leq \epsilon$$

where i is the sample natural mutual information

$$i\left(\underline{x}^n; \underline{y}^n\right) = \ln\left(\frac{P_{\underline{Y}^n|\underline{X}^n}(\underline{y}^n|\underline{x}^n)}{P_{\underline{Y}^n}(\underline{y}^n)}\right)$$

For every n and each code in the ensemble, the decoder, given $\underline{y}^n$, selects the message m for which $\left(\underline{x}^n(m), \underline{y}^n\right) \in T_\epsilon^n$.

We assume an error if there are no such codewords or more than one codeword.

Upper bound on probability

Let λ_m be the event that, given message m enters the system, an error occurs.

The mean probability of error over all ensemble of codes is

$$E[\lambda_m] = P(\lambda_m = 1)$$

(indicator function)

Error occurs when

$$\left(\underline{x}^n(m), \underline{y}^n\right) \notin T_\epsilon^n$$

or

$$\left(\underline{x}^n(m'), \underline{y}^n\right) \in T_\epsilon^n \text{ for } m' \neq m$$

Upper bound on probability

Hence, through the union bound

$$\begin{aligned} E[\lambda_m] &= P\left(\left(\left(\underline{x}^n(m), \underline{y}^n\right) \notin T_\epsilon^n\right)\right. \\ &\quad \left. \cup \bigcup_{m' \neq m} \left(\underline{x}^n(m'), \underline{y}^n\right) \in T_\epsilon^n | m\right) \\ &\leq P\left(\left(\underline{x}^n(m), \underline{y}^n\right) \notin T_\epsilon^n\right) \\ &+ \sum_{m' \neq m} P\left(\left(\underline{x}^n(m'), \underline{y}^n\right) \in T_\epsilon^n | m\right) \end{aligned}$$

Bound on the pair not being typical

The probability of the pair not being typical approaches 0 as $n \to \infty$

$$\frac{i\left(\underline{x}^n(m);\underline{y}^n\right)}{n} = \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{P_{Y|X}(y_i|x_i)}{P_Y(y_i)}\right)$$

Through the WLLN, the above converges in probability to

$$C = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_X(x)P_{Y|X}(y|x)\ln\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right)$$

Hence, for any ϵ

$$\lim_{n\to\infty} P\left(\left(\underline{x}^n(m),\underline{y}^n\right) \in T_\epsilon^n|m\right) \to 0$$

Upper bound to probability of several typical sequences

Consider $m' \neq m$

Given how the codebook was chosen, the variables $\underline{X}^n(m'), \underline{Y}^n$ are independent *conditioned on m having been transmitted*

Hence

$$P\left((\underline{X}^n(m'), \underline{Y}^n) \in T^n_\epsilon | m\right)$$

$$= \sum_{\left(\underline{x}^n(m'), \underline{y}^n\right) \in T^n_\epsilon} P_{\underline{X}^n}(\underline{x}^n(m')) P_{\underline{Y}^n}(\underline{y}^n)$$

Because of the definition of T^n_ϵ, for all pairs in the set

$$P_{\underline{Y}^n}(\underline{y}^n) \leq P_{\underline{Y}^n|\underline{X}^n}\left(\underline{y}^n|\underline{x}^n(m')\right) e^{-n(C-\epsilon)}$$

Upper bound to probability of several typical sequences

$$
\begin{aligned}
& P\left(\underline{X}^n(m'), \underline{Y}^n \in T_\epsilon^n | m\right) \\
\leq & \sum_{\left(\underline{x}^n(m'), \underline{y}^n\right) \in T_\epsilon^n} P_{\underline{Y}^n|\underline{X}^n}\left(\underline{y}^n | \underline{x}^n(m')\right) e^{-n(C-\epsilon)} \\
& P_{\underline{X}^n}(\underline{x}^n(m')) \\
\leq & e^{-n(C-\epsilon)}
\end{aligned}
$$

Weak coding theorem

$E[\lambda_m]$ is upper bounded by two terms that go to 0 as $n \to \infty$

thus the average probability of error given that m was transmitted goes to 0 as $n \to \infty$

This is the average probability of error averaged over the ensemble of codes, therefore $\forall \delta > 0$ and for any rate $R < C$ there must exist a code length n with average probability or error less than δ

Thus, we can create a sequence of codes with maximal probability of error converging to 0 as $n \to \infty$

Weak coding theorem

How do we make codebooks that are "good"?

Clearly, some of the codebooks are very bad, for instance codebooks in which all the codewords are identical

Even within a more reasonable codebook, some codewords may do very badly

Expurgated codes. Let us introduce a r.v. M uniformly distributed over all possible values of m

The average probability of error over all codebooks is $E_M[E_{codebooks}[\lambda_M]]$

Let us select n large enough $E[\lambda_m] \leq \delta$ for every m, hence

$$E_M[E_{codebooks}[\lambda_M]] \leq \delta$$

so

$$E_{codebooks}[E_M[\lambda_M]] \leq \delta$$

Weak coding theorem

Let us introduce a r.v. S uniformly distributed over all possible values of s

$$E_S[E_M[\lambda_M^S]] \le \delta$$

Markov inequality states that

$$P(E_M[\lambda_M^S] \ge 2E_S[E_M[\lambda_M^s]]) \le \frac{1}{2}$$

since we picked the messages to have uniform distribution, we have that the half of the messages with lowest probability of error have probability of error $\le 2\delta$

Weak coding theorem

We can therefore create a codebook using only the best half codewords

What is the penalty?

The rate is reduced because $\lfloor \frac{M}{2} \rfloor = \lfloor e^{nR} \rfloor$

so $R \approx \frac{\ln(M)}{n} - \frac{\ln 2}{n}$

which is arbitrarily close to $\frac{\ln(M)}{n}$ as $n \to \infty$

(M grows with n as needed to maintain rate)

Interpretation

Random codebooks are good!

How can we implement coding strategies that are random and how well do they perform?

How well do random codebooks perform for finite length codewords?