Lecture 27

Instructor: Himanshu Tyagi
Scribe: Dhruv Jawali
19-Oct-2015

Contents
1 Large Probability Sets 1
2 Random Code Construction 2
3 Alternative Proof of the Random Coding Argument 2
4 Gaussian Channels 4

1 Large Probability Sets

We can prove the random coding argument using large probability sets, as we shall see. Define a set $A_\delta$ as follows:

$$A_\delta = \left\{ (x, y) \in X^n \times Y^n \left| \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_n(y_i \mid x_i)}{Q(y_i)} \geq I(P; W) - \delta \right. \right\}$$

Claim: $\lim_{n \to \infty} P_{X^n Y^n}(A_\delta) = 1$

Proof: Let $Z_1, ..., Z_n$ be I.I.D with common distribution $P_Z$.

$$\lim_{n \to \infty} \Pr \left( \left| E[z] - \frac{1}{n} \sum_{i=1}^{n} Z_i \right| \leq \delta \right) \leftrightarrow \lim_{n \to \infty} \Pr \left( \frac{1}{n} \sum_{i=1}^{n} Z_i \geq E[z] - \delta \right) = 1;$$

By weak law of large numbers. Putting $Z_i = \log \frac{W_n(y_i \mid x_i)}{Q(y_i)},$

$$\lim_{n \to \infty} \Pr \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_n(y_i \mid x_i)}{Q(y_i)} \geq I(P; W) - \delta \right) = 1;$$

And hence, the result.
2 Random Code Construction

Codebook Generation:
The code book $X$, containing $N$ codewords of length $n$ is randomly generated by a distribution $P_X$ over $\mathcal{X}$ where $X_{ij}$ are I.I.D.

$$X = \begin{pmatrix} X_{11} & \ldots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{N1} & \ldots & X_{Nn} \end{pmatrix}_{N \times n}$$

Decoding:
Given channel output $y$, let $i(y)$ be the unique index $i$ such that $(x_i, y) \in A_\delta$.

We shall show that the average probability of error goes to zero, as long as the rate $R < I(P; W)$ i.e.

$$E_P[\epsilon(x_i)] \to 0 \text{ as } n \to \infty$$

where $\epsilon(x_i)$ refers to the average probability of error of the $i^{th}$ codeword.

3 Alternative Proof of the Random Coding Argument

$$E_P[\epsilon(x_i)] = \sum_{\bar{x} \in \mathcal{X}} \Pr(\bar{x}) \epsilon(\bar{x})$$

$$= \sum_{\bar{x} \in \mathcal{X}} \Pr(\bar{x}) \left[ \frac{1}{N} \sum_{i=1}^{N} \Pr(\text{error} \mid x_i \text{ was sent, } X \text{ was used}) \right]$$

$$= \frac{1}{N} \sum_{i=1}^{N} E[\Pr(\text{error} \mid x_i \text{ was sent, } X \text{ was used})]$$

$$= E[\Pr(\text{error} \mid x_1 \text{ was sent, } X \text{ was used})]$$

Where the last step follows because $x_i$ are I.I.D and it does not matter which $i$ we choose. There are two types of error events associated with this scheme:

$$\mathcal{E}_1 = \{(\bar{x}, y) \in \mathcal{X}^n \times \mathcal{Y}^n \mid (x_i, y) \notin A_\delta \}$$

$$\mathcal{E}_2 = \{(\bar{x}, y) \in \mathcal{X}^n \times \mathcal{Y}^n \mid (\bar{x}, y) \in A_\delta, i \neq 1 \}$$
\[ E_p[\epsilon(x_i)] = \sum_{x \in X^n} P^n_x(x) E[\Pr(\text{error} | x_i \text{ was sent, } X \text{ was used}) | x_1 = x] \] (5)

\[ \leq \sum_{x \in X^n} P^n_x(x) E \left[ \sum_{y : (x, y) \in E_1} W^n(y | x) + \sum_{y : (x, y) \in E_2} W^n(y | x) \bigg| x_1 = x \right] \] (6)

\[ = \sum_{x \in X^n} P^n_x(x) \sum_{y : (x, y) \in E_1} W^n(y | x) \] (7)

\[ + \sum_{x \in X^n} P^n_x(x) E_{x_2 \ldots x_N} \left[ \sum_{y : (x, y) \in E_2} W^n(y | x) \right] \] (8)

We now consider the two terms of error, (7) and (8) separately.

\[ (7) = \sum_{x \in X^n} P^n_x(x) \sum_{y : (x, y) \in E_1} W^n(y | x) \] (9)

\[ = \sum_{(x, y) \notin A_0} P^n_x(x) W^n(y | x) \] (10)

\[ = P^n_{XY}(A_0) \] (11)

\[ \to 0 \text{ as } n \to \infty \text{ from the initial claim.} \] (12)

\[ (8) \leq \sum_{x \in X^n} P^n_x(x) \sum_{i=2}^N E \left[ W^n(\{(x, y) \in A_0\} | x) \right] \] (13)

\[ = \sum_{x \in X^n} P^n_x(x) \sum_{i=2}^N E \left[ \sum_{y : (x, y) \in A_0} W^n(y | x) \right] \] (14)

\[ = \sum_{i=2}^N E \left[ \sum_{y : (x, y) \in A_0} \sum_{x \notin X} P^n(x) W^n(y | x) \right] \] (15)

\[ = \sum_{i=2}^N E \left[ \sum_{y : (x, y) \in A_0} Q^n(y) \right] \] (16)

Applying \( E[f(X)] = \sum_{x \in X} P(x)f(x) \) to the summand,

\[ = \sum_{i=2}^N \sum_{x \in X} P^n(x) \sum_{y : (x, y) \in A_0} Q^n(y) \] (17)

\[ = \sum_{i=2}^N \sum_{(x, y) \in A_0} P^n(x) Q^n(y) \] (18)

\[ = \sum_{i=2}^N \sum_{(x, y) \in A_0} P^n(x) Q^n(y) \] (19)
Recall: $\forall (x, y) \in A_\delta,$

\[
\frac{1}{n} \log \frac{W^n(y \mid x)P^n(x)}{Q^n(y)P^n(x)} \geq I(P; W) - \delta
\]

\[
\Rightarrow P^n(x)Q^n(y) \leq 2^{-n(I(P; W) - \delta)} P^n(x)W^n(y \mid x)
\]

Plugging this back into (19),

\[
\sum_{i=2}^{N} E \left[ \sum_{(x_i, y) \in A_\delta} Q^n(y) \right] \leq \sum_{i=2}^{N} 2^{-n(I(P; W) - \delta)} \sum_{(x_i, y) \in A_\delta} P^n(x)W^n(y \mid x) (20)
\]

\[
\leq \sum_{i=2}^{N} 2^{-n(I(P; W) - \delta)} (21)
\]

\[
\leq N 2^{-n(I(P; W) - \delta)} (22)
\]

\[
= 2^{-n(I(P; W) - R - \delta)} (23)
\]

\[
\rightarrow 0 \text{ as } n \rightarrow 0 \text{ as long as } R < I(P; W) (24)
\]

Thus, a random coding scheme is good enough to achieve capacity.

4 Gaussian Channels

We now consider Additive "White" Gaussian Noise Channels. These channels have inputs from the real number set. We will now have to deal with probability density functions rather than mass functions.

Such a channel $W : \mathcal{X} \rightarrow \mathcal{Y}$ is given by the conditional density function $f_{Y \mid X}(y \mid x).$ E.g:

\[
W(Y \in (-a, a) \mid X = x) = \int_{-a}^{a} f_{Y \mid X}(y \mid x) dy
\]

A code of length $n$ and size $2^k$ consists of:

An encoder $f : \{0, 1\}^k \rightarrow \mathcal{X}^n$

and a decoder $\varphi : \mathcal{Y}^n \rightarrow \{0, 1\}^k \cup \{\phi\}$
E.g: Suppose $Y = x + N$;

$$N \sim \mathcal{N}(0, \sigma^2);$$

$$f_X(i) = iM; -\infty \leq i \leq \infty, M \in \mathcal{R};$$

In the above example, we can suggest codes of infinite rates as there is no constraint on $i$. In practical scenarios, we are limited by the power that a signal can carry, a power constraint.

Suppose average power $\leq p$; $\implies$ We can use $x = (x_1, \ldots, x_n)$ s.t. $\frac{1}{n} \sum_{i=1}^{n} x_i^2 \leq p$.

Using the sphere packing argument, number of codewords $= \left( \frac{p + \sigma^2}{\sigma^2} \right)^{n/2} = 2^n \left[ \log \left( 1 + \frac{p}{\sigma^2} \right) \right]$