LECTURE 10

Last time:

- Joint AEP
- Coding Theorem

Lecture outline

- Error Exponents
- Strong Coding Theorem

Reading: Gallager, Chapter 5.

Review

- Joint AEP

$$\begin{array}{cccc} A_\epsilon^{(n)}(X) & \times A_\epsilon^{(n)}(Y) & \text{vs.} & A_\epsilon^{(n)}(X,Y) \\ 2^{nH(X)} & \times 2^{nH(Y)} & >> & 2^{nH(X,Y)} \end{array}$$

  - Passing $\underline{x}^n$ through the channel to obtain $\underline{y}^n$, $(\underline{x}^n, \underline{y}^n)$ are jointly typical with high probability.
  - For another independently chosen $\underline{\tilde{x}}^n$, $(\underline{\tilde{x}}^n, \underline{y}^n)$ are jointly typical with probability $2^{nI(X;Y)}$.
- Coding Theorem
  - Random coding
  - Joint typicality decoding
  - Converse proved by using Fano's inequality.
- A Possible Confusion: i.i.d. input distribution vs. transmitting independent symbols.

Remaining Topics

- Can we get rid of the random coding? Instead, we will get a closer look of random coding.
- Joint typicality decoding vs. Maximum Likelihood decoding.

Example: Binary source sequence passing through BSC.

- Finite codeword length n.

Our plan for the next two lectures

- Maximum likelihood decoding.
- Upper bound of the error probability.
- Random coding error exponent.
- Binary symmetric channel as an example.

Maximum Likelihood Decoding

Notations

- Message W uniformly distributed in $\{1, 2, \ldots, M\}$. $M = 2^{nR}$.
- Encoder transmit codeword $\underline{x}^n(m)$ if the incoming message is $W = m$.
- Receiver receives $\underline{y}^n$, and find the most likely transmitted codeword.

$$\hat{W} = \arg\max_m P_{\underline{Y}^n|\underline{X}^n}(\underline{y}^n|\underline{x}^n(m))$$

- Define $\mathcal{Y}_m$ as the set of $\underline{y}^n$'s that decodes to m.
- The probability of error conditioned on $W = m$ is transmitted:

$$P_{e,m} = \sum_{\underline{y}^n \in \mathcal{Y}_m^c} P_{\underline{Y}^n|\underline{X}^n}(\underline{y}^n|\underline{x}^n(m))$$

Pairwise Error Probability

Consider the case $M = 2$.

$$P_{e,1} = \sum_{\underline{y}^n \in \mathcal{Y}_1^c} P(\underline{y}^n | \underline{x}^n(1))$$

- We really hate the $\mathcal{Y}_1^c$, since we have to figure out the decision region. Can we sum over the entire set $\mathcal{Y}^n$ without dealing with the discontinuity?
- Consider any $\underline{y}^n \in \mathcal{Y}_1^c$, by definition

  $$P(\underline{y}^n | \underline{x}^n(2)) \geq P(\underline{y}^n | \underline{x}^n(1)),$$

  so we can bound

  $$\begin{aligned} P_{e,1} &\leq \sum_{\underline{y}^n \in \mathcal{Y}_1^c} P(\underline{y}^n | \underline{x}^n(1)) \left[\frac{P(\underline{y}^n | \underline{x}^n(2))}{P(\underline{y}^n | \underline{x}^n(1))} \right]^s \\ &= \sum_{\underline{y}^n \in \mathcal{Y}_1^c} P(\underline{y}^n | \underline{x}^n(1))^{1-s} P(\underline{y}^n | \underline{x}(2))^s \\ &\leq \sum_{\underline{y}^n} P(\underline{y}^n | \underline{x}^n(1))^{1-s} P(\underline{y}^n | \underline{x}(2))^s \end{aligned}$$

  for any $s \in (0, 1)$.

Random Codewords

Choose the codeword $\underline{x}^n(1)$ and $\underline{x}^n(2)$ i.i.d. from the distribution P_X (or equivalently $P_{\underline{X}^n}$)

$$\begin{aligned}\overline{P}_{e,1} &= \sum_{\underline{x}^n(1)} P_{\underline{X}^n}(\underline{x}^n(1)) \sum_{\underline{y}^n} P(\underline{y}^n|\underline{x}^n(1)) \\ &\quad \times P(\text{error}|W=1,\underline{x}^n(1),\underline{y}^n)\end{aligned}$$

and

$$\begin{aligned}& P(\text{error}|W=1,\underline{x}^n(1),\underline{y}^n) \\ \leq & \quad \sum_{\underline{x}^n(2)} P_{\underline{X}^n}(\underline{x}^n(2)) \left[\frac{P(\underline{y}^n|\underline{x}^n(2))}{P(\underline{y}^n|\underline{x}^n(1))}\right]^s\end{aligned}$$

- In general, this is a good bound for both the fixed and the random codewords.
- Random coding allows for generalization to many codewords. The upper bound allows us to compute the error probability without dealing with specific decision regions.

Example: Binary source/BSC

Let $\underline{x}^n(1)$ and $\underline{x}^n(2)$ be the all 0 and the all 1 words.

- Use DMC, we have for $m = 1, 2$, and any $s \in (0, 1)$,

$$
\begin{aligned}
P_{e,m} &\leq \sum_{y_1} \sum_{y_2} \cdots \sum_{y_n} \prod_{i=1}^{n} P(y_i|x_{1,i})^{1-s} P(y_i|x_{2,i})^s \\
&= \prod_{i=1}^{n} \sum_{y_i} P(y_i|x_{1,i})^{1-s} P(y_i|x_{2,i})^s
\end{aligned}
$$

Example: Binary source/BSC

Plug in the specific choices of codewords:

$$\sum_{y_i} P(y_i|x_{1,i})^{1-s} P(y_i|x_{2,i})^s =$$
$$\epsilon^{1-s}(1-\epsilon)^s + \epsilon^s(1-\epsilon)^{1-s}$$

Optimize to get $s^* = 1/2$, and

$$P_{e,m} \leq [2\sqrt{\epsilon(1-\epsilon)}]^N$$

Alternative Approach

Condition on the all 0 word is transmitted, error occurs when there are more than $n/2$ 1's,

$$P_{e,1} \approx 2^{-nD(\frac{1}{2}||\epsilon)} = 2^{n[\log 2 + \frac{1}{2}\log\epsilon + \frac{1}{2}\log(1-\epsilon)]}$$

- The upper bound is quite tight!
- Similar development can be done with random codewords.
- We are now one step away from the case with many codewords.

BSC with Many Codewords

- Can we generalize this to many codeword by using the union bound? Assume there are 2^{nR} codewords,

$$\text{union bound of } P_{e,m} = \sum_{m' \neq m} P(m \to m') \geq 2^{nR} 2^{-nD(\frac{1}{2}||\epsilon)}$$

The error probability goes to 0 if

$$R - D\left(\frac{1}{2}||\epsilon\right) < 0$$

- This says the probability of error decays exponentially with n.

The Error Exponent

Rewrite

$$P_{e,m,union} = 2^{-n(D(1/2||\epsilon)-R)}$$

- The error exponent is $E_u(R) = D\left(\frac{1}{2}||\epsilon\right) - R$.
- As long as R is small enough such that $E_u(R) > 0$, the error probability decays with n exponentially.
- **Question** Does this give the capacity?
- **Too bad!**, the maximum data rate is not the capacity.

$$\begin{aligned}
& 1 - H(\epsilon) - D\left(\frac{1}{2}||\epsilon\right) \\
= \; & \log 2 + \epsilon \log \epsilon + (1-\epsilon)\log(1-\epsilon) \\
& -[\log 2 + \frac{1}{2}\log \epsilon + \frac{1}{2}\log(1-\epsilon)] \\
= \; & \left(\frac{1}{2} - \epsilon\right)\log\frac{1-\epsilon}{\epsilon} \\
> \; & 0
\end{aligned}$$

A Better Way than the Union Bound

Lemma For any $\rho \in (0, 1]$,

$$P\left(\bigcup_m A_m\right) \leq \left[\sum_m P(A_m)\right]^\rho$$

Proof

$$P\left(\bigcup_m A_m\right) \leq \begin{cases} \sum_m P(A_m) \\ 1 \end{cases}$$

Idea: use ρ to compensate serious overlap with a cost.

Now define event

$$A_{m'} = \{m \to m' | W = m, \underline{x}^n(m), \underline{y}^n\}$$

and we have just computed

$$P(A_{m'}) \leq \sum_{\underline{x}^n(m')} P_{\underline{X}^n}(\underline{x}^n(m')) \frac{P(\underline{y}^n | \underline{x}^n(m'))^s}{P(\underline{y}^n | \underline{x}^n(m))^s}$$

Upper Bound for the Error Probability

$$
\begin{aligned}
& P(\text{error}|W = m, \underline{x}^n(m), \underline{y}^n) \\
= \;& P\left(\bigcup_{m' \neq m} A'_m\right) \\
\leq \;& (M-1)^\rho \left[\sum_{\underline{x}^n(m')} P_{\underline{X}^n}(\underline{x}^n(m')) \frac{P(\underline{y}^n|\underline{x}^n(m'))^s}{P(\underline{y}^n|\underline{x}^n(m))^s}\right]^\rho
\end{aligned}
$$

Average over $\underline{x}^n(m), \underline{y}^n$,

$$
\begin{aligned}
& \overline{P}_{e,m} \\
\leq \;& \sum_{\underline{y}^n} \sum_{\underline{x}^n(m)} P_{\underline{X}^n}(\underline{x}^n(m)) P(\underline{y}^n|\underline{x}^n(m))^{1-s\rho} \\
& (M-1)^\rho \left[\sum_{\underline{x}^n(m')} P_{\underline{X}^n}(\underline{x}^n(m')) P(\underline{y}^n|\underline{x}^n(m'))^s\right]^\rho
\end{aligned}
$$

for any $s, \rho \in (0, 1]$.

take $s = 1/(1+\rho)$.

Upper Bound

Theorem

$$\overline{P}_{e,m} \leq (M-1)^{\rho} \sum_{\underline{y}^n} \left[\sum_{\underline{x}^n} P_{\underline{X}^n}(\underline{x}^n) P(\underline{y}^n|\underline{x}^n)^{\frac{1}{1+\rho}} \right]^{(1+\rho)}$$

Corollary Apply DMC

$$\begin{aligned} & \overline{P}_{e,m} \\ \leq \; & (M-1)^{\rho} \\ & \prod_{i=1}^{n} \left[\sum_{y_i} \left(\sum_{x_i} P_X(x_i) P_{Y|X}(y_i|x_i)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right] \\ = \; & (M-1)^{\rho} \left[\sum_{y} \left(\sum_{x} P_X(x) P_{Y|X}(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right]^{n} \end{aligned}$$

Random Coding Error Exponent

For a fixed input distribution P_X, define

$$E_0(\rho) = -\log \left[\sum_y \left(\sum_x P_X(x) P_{Y|X}(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right]$$

Then the average probability of error

$$\overline{P}_{e,m} \leq 2^{-n(E_0(\rho) - \rho R)}$$

As long as the **random coding error exponent**

$$E_r(R) = \max_{\rho \in [0,1]} [E_0(\rho) - \rho R]$$

is positive, the error probability can be driven to 0 as $n \to \infty$.

The Behavior of the Error Exponent

Facts:

- $E_0(\rho) \geq 0$ with equality only at $\rho = 0$.
- $\frac{\partial E_0(\rho)}{\partial \rho} \geq 0$.
- $\frac{\partial E_0(\rho)}{\partial \rho} \leq I(X;Y)$, with equality at $\rho = 0$.
- $E_0(\rho)$ is concave in ρ.

Consider

$$E_r(R) = \max_{\rho \in [0,1]} [E_0(\rho) - \rho(R)]$$

- Ignore the constraint, the maximum occurs at $R = \partial E_0(\rho)/\partial \rho \mid_{\rho^*}$.
- The maximizing ρ^* lies in $[0, 1]$ if

$$\left.\frac{\partial E_0(\rho)}{\partial \rho}\right|_{\rho=1} \leq R \leq \left.\frac{\partial E_0(\rho)}{\partial \rho}\right|_{\rho=0} = I(X;Y)$$

- For any $R < I(X;Y)$, we get positive error exponent, and the error probability can be driven to 0 as $n \to \infty$.

Summary

- We have proved the coding theorem in another way.
- For $R < C$, the error probability decays exponentially with n.

Remaining Questions

- Is this a good bound?
- We have chosen the random codes and computed the average performance. Is there any specific code that can do better than this?
- The two pieces of the error exponent curve is mysterious.