LECTURE 11

Last time:

- Error Exponents
- Strong Coding Theorem

Lecture outline

- Binary source/BSC.
- Typical error events.

Review

- Strong Coding Theorem,

$$\overline{P}_{e,m} \leq \exp[-nE_r(R)]$$

where

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

and

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

- $E_r(R) > 0$ for any $R < C$.

- For any R, the maximizing ρ is the slope of the $E_r(R) \sim R$ curve at R.

- **Definition** The critical rate

$$R_{crit} = \left. \frac{\partial E_0(\rho)}{\partial \rho} \right|_{\rho=1}$$

- The maximizing $\rho \in [0,1]$ if $R_{crit} \leq R \leq I(X;Y)$.

- For $R < R_{crit}$, the slope of $E_r(R) \sim R$ is -1.

A Complete Picture of the Reliability Function

- To improve the random coding bound, expurgate bad codes
- For a lower bound of the error probability: sphere packing bound.

Conclusion

- For $R < C$, error probability decays with n exponentially.
- For $R > R_{crit}$, random codes are optimal, the average error probability achieves the highest possible error exponent.
- For $R > R_{crit}$, the union bound is not tight, there is no one dominating pair-wise error event.
- For $R < R_{crit}$, the union bound is fine, but random coding is not optimal.

Example: Binary Source/BSC

Choose the input to be equiprobable. Define

$$\tau = \frac{\sqrt{\epsilon}}{\sqrt{\epsilon} + \sqrt{1-\epsilon}}$$

- For $R \geq \log 2 - H(\tau) = D(\tau || \frac{1}{2})$,

$$\begin{aligned} R &= D\left(\gamma || \frac{1}{2}\right) \\ E_r(R) &= D(\gamma || \epsilon) \end{aligned}$$

  for $\gamma \in (\epsilon, \tau)$.

- For $R < D(\tau || \frac{1}{2})$,

$$E_r(R) = \log 2 - \log(1 + 2\sqrt{\epsilon(1-\epsilon)}) - R$$

Gallager: *The most significant point about this example is that, even for such a simple channel, there is no simple way to express* $E_r(R)$ *except in parametric form.*

A Little Large Deviation Theory

Suppose $h(x)$ is bounded and continuous on $[0,1]$,

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^1 \exp[-nh(x)]dx = -\min_{x\in[0,1]} h(x)$$

Chernoff exponent Let $\underline{X}^n$ be a sequence of $Bern(p)$ r.v.s, and $w(\underline{X}^n)$ be the hamming weight of the vector, for $\tau > p$,

$$P(w(\underline{X}^n) \geq N\tau) \doteq 2^{-nD(\tau||p)}$$

Proof

Denote by E_q the event that a $Bern(p)$ sequence $\underline{X}^n$ is typical w.r.t. another distribution q

Recall

$$P(E_q) \doteq 2^{-nD(q||p)}.$$

For large enough n, the probability $P(\cup_{q\geq\tau}E_q)$ is dominated by $P(\cup_{q\in[\tau,\tau+\epsilon)}E_q)$ for an arbitrarily small ϵ.

Output Centered Analysis

For random codes on the BSC:

Assume $\underline{X}^n(0)$ is transmitted, and $\underline{Y}^n$ is observed. Let the other codewords be $\underline{X}^n(i), i = 1, \ldots, M-1$.

The joint distribution is

$$\begin{aligned} &P\left(\underline{X}^n(0), \underline{Y}^n, \{\underline{X}^n(i), i = 1, \ldots M-1\}\right) \\ &\qquad = P(\underline{X}^n(0), \underline{Y}^n) \prod_i P(\underline{X}^n(i)) \end{aligned}$$

Forney: We are only interested in the distances between the codewords and the output $\underline{Y}^n$.

Consider this as two subsystems.

- Translate the correct codeword to the noise vector

$$\underline{\boldsymbol{\Delta}} = \underline{X}^n(0) \oplus \underline{Y}^n$$

$\underline{\underline{\Delta}}$ has i.i.d. $\{\epsilon, 1-\epsilon\}$ entries.

- Translate the other codewords to

$$\underline{z}_i = \underline{X}^n(i) \oplus \underline{Y}^n$$

For i, $\underline{z}_i$ has equiprobable entries.

Now the error occurs if $w(\underline{\underline{\Delta}}) > w(\underline{z}_i)$ for some $i = 1, \ldots, M-1$.

We compute the error probability and ask the question: is the error caused by

- large noise vector?
- or some incorrect codeword being too close?

The Exponents

- For the noise vector,

$$P(w(\underline{\Delta}) \geq n\gamma) \doteq 2^{-nE_I}$$

where

$$E_I = \begin{cases} D(\gamma||\epsilon) & \gamma > \epsilon \\ 0 & \gamma \leq \epsilon \end{cases}$$

- For the incorrect codewords,

$$P(w(\underline{z}_i) \leq n\gamma) \doteq 2^{nD(\gamma||\frac{1}{2})}$$

Now

$$\begin{aligned} P(\min_i w(\underline{z}_i) \leq n\gamma) &= P\left(\bigcup_i \{w(\underline{z}_i) \leq n\gamma\}\right) \\ &\doteq 2^{-nE_{II}} \end{aligned}$$

where

$$E_{II} = \begin{cases} D(\gamma||\frac{1}{2}) - R, & D(\gamma||\frac{1}{2}) \geq R \\ 0 & D(\gamma||\frac{1}{2}) \leq R \end{cases}$$

For a given R, let γ_R^* satisfy $D(\gamma_R^*||\frac{1}{2}) = R$

- For any $\gamma > \gamma_R^*$, or equivalently $R \geq D(\gamma || \frac{1}{2})$, there are exponentially many codewords with $w(\underline{z}_i) < n\gamma$.
- For any $\gamma < \gamma_R^*$, or equivalently $R \leq D(\gamma || \frac{1}{2})$, the probability that there exist a $\underline{z}_i$ with $w(\underline{z}_i) < n\gamma$, is exponentially small.
- γ_R^* is the typical min distance at rate R, also called Gilbert-Varshamov distance.

Channel Capacity

- If $R < D(\epsilon||\frac{1}{2})$, or equivalently, $\gamma_R^* > \epsilon$, then we can find $\gamma > \epsilon$ such that $D(\gamma||\frac{1}{2}) \geq R$.
  - Decoder decodes if $\exists!$ codeword that is within $n\gamma$ distance from $\underline{y}^n$, and claim error otherwise.
  - The probability of both $\{w(\underline{\Delta}) > n\gamma\}$ and $\{\min_i w(\underline{z}_i) < n\gamma\}$ are exponentially small, so the decoding error probability is exponentially small.

- If $R > D(\epsilon||\frac{1}{2})$, then for γ such that $R > D(\gamma||\frac{1}{2}) \geq D(\epsilon||\frac{1}{2})$, both $\{w(\underline{\Delta}) > n\gamma\}$ and $\{\min_i w(\underline{z}_i) < n\gamma\}$ occurs with probability $\doteq 1$, so the rate is not supported.

- Notice $C = \log_2 - H(\epsilon) = D(\epsilon||\frac{1}{2})$.

Error Exponent

Suppose that $R < D(\epsilon||\frac{1}{2})$, we now find the error exponent for $P_e = P(w(\underline{\Delta}) \geq \min_i w(\underline{z}_i))$.

Define type γ error as the event

$$\mathcal{E}_\gamma = \{w(\underline{\Delta}) \geq n\gamma\} \cap \{\min_i w(\underline{z}_i) \leq n\gamma\}$$

and $P(\mathcal{E}_\gamma) \doteq 2^{-nE_\gamma}$.

Now

$$\begin{aligned} E_\gamma &= E_I + E_{II} \\ &= \begin{cases} D(\gamma||\epsilon) + D(\gamma||\frac{1}{2}) - R & D(\gamma||\frac{1}{2}) \geq R, \gamma > \epsilon \\ D(\gamma||\epsilon) & D(\gamma||\frac{1}{2}) \leq R \end{cases} \end{aligned}$$

The error exponent is

$$E_r(R) = \min_\gamma E_\gamma$$

Error Exponent

- The condition $D(\gamma||\frac{1}{2}) \geq R, \gamma > \epsilon$ is equivalent to $\epsilon < \gamma \leq \gamma_R^*$.
- The optimum must occur at $\gamma \leq \gamma_R^*$.

$$E_r(R) = \min \gamma \in (\epsilon, \gamma_R^*] \left[D(\gamma||\epsilon) + D(\gamma||\frac{1}{2}) - R \right]$$

- First ignore the constraint, minimum occurs at $\gamma = \tau$, where $D(\gamma||\epsilon) + D(\gamma||\frac{1}{2})$ is minimized. Can solve to have

$$\tau = \frac{\sqrt{\epsilon}}{\sqrt{\epsilon} + \sqrt{1-\epsilon}}$$

Define $R_{crit} = D(\tau||\frac{1}{2})$. The minimum is

$$\begin{aligned} E_0 &= D(\tau||\epsilon) + D(\tau||\frac{1}{2}) \\ &= \log 2 - \log(1 + 2\sqrt{\epsilon(1-\epsilon)}) \end{aligned}$$

- If $\tau < \gamma_R^*$, or equivalently $R < R_{crit}$, the minimum is achieved at $\gamma = \tau$,

$$E_r(R) = E_0 - R$$

- If $\tau > \gamma_R^*$, or equivalently $R > R_{crit}$, the minimum occurs at $\gamma = \gamma_R^*$,

$$\begin{aligned} R &= D(\gamma_R^* || \frac{1}{2}) \\ E_r(R) &= D(\gamma_R^* || \epsilon) \end{aligned}$$

Discussions

Main Conclusion The error mechanisms are different in the high rate regime $R_{crit} \leq R < C$, and the low rate regime $R < R_{crit}$.

- In the high rate regime, error occurs when the noise is so large it reaches γ_R^*.
  - Confusion occurs among exponentially many codewords.
  - Union bound is not tight.
  - Draw a sphere of radius γ_R^* around each codeword, as long as the $\underline{y}^n$ lies in the sphere, error does not occur – sphere packing argument.
  - Cannot improve by expurgating bad codewords, since there are too many of them.

- In the low rate regime, error occurs when the noise is within the sphere of radius γ_R^*, but some atypically bad codeword $\underline{X}^n(i)$ is too close to $\underline{y}^n$.
  - Error occurs at one particular bad codeword.
  - Union bound is fine: $P_e \doteq Me^{nE_0}$.
  - Error is caused by atypically bad codes from the ensemble. Can improve by expurgating the bad codeword.