EE 231A Information Theory
Lecture 9
Converse to the Channel Coding Theorem

- Fano's Inequality
- Considering possibly dependent inputs
- Proof of the converse

Part A: Fano's Inequality

Fano's inequality

- Stated in section 7.9, but proof in section 2.10.
- For a DMC and a $(2^{nR}, n)$ code with input messages uniformly distributed.

$$H(W \mid \hat{W}) \le 1 + P(\hat{W} \ne W) nR$$

An upper bound on $H(W \mid \hat{W})$

$$\boxed{E = \begin{cases} 1 & \text{if } \hat{W} \ne W \\ 0 & \text{if } \hat{W} = W \end{cases}}$$

$$\begin{aligned} H(E, W \mid \hat{W}) &= H(W \mid \hat{W}) + H(E \mid W, \hat{W}) \\ &= H(E \mid \hat{W}) + H(W \mid E, \hat{W}) \end{aligned}$$

$$\begin{aligned} H(W \mid \hat{W}) &= \underbrace{H(E \mid \hat{W})}_{\le 1} + H(W \mid E, \hat{W}) - \cancelto{0}{H(E \mid W, \hat{W})} \\ &\le 1 + H(W \mid E, \hat{W}) \end{aligned}$$

$$\boxed{H(W \mid \hat{W}) \le 1 + H(W \mid E, \hat{W})}$$

An upper bound on $H(W \mid E, \hat{W})$

$$
\begin{aligned}
H(W \mid E,\hat{W}) &= P(E=0)\cancelto{0}{H(W \mid \hat{W}, E=0)} \\
&\quad + P(E=1)H(W \mid \hat{W}, E=1) \\
&= P(\hat{W} \neq W)H(W \mid \hat{W}, E=1) \\
&\leq P(\hat{W} \neq W)\log|\mathcal{W}| \\
&= P(\hat{W} \neq W)nR
\end{aligned}
$$

Thus $\boxed{H(W \mid \hat{W}) \leq 1 + P(\hat{W} \neq W)nR}$

Part B: Considering Possibly Dependent Inputs

Lemma 7.9.2

Let Y^n be the result of passing X^n through a discrete memoryless channel of capacity C.

Then $I\left(X^n;Y^n\right) \le nC$ for all $p\left(x^n\right)$.

Proof of Lemma 7.9.2

$$
\begin{aligned}
I\left(X^n;Y^n\right) &= H\left(Y^n\right) - H\left(Y^n \mid X^n\right) \\
&= H\left(Y^n\right) - \sum_{i=1}^{n} H\left(Y_i \mid Y_1,\cdots,Y_{i-1,}X^n\right) \\
&= H\left(Y^n\right) - \sum_{i=1}^{n} H\left(Y_i \mid X_i\right) \\
&\le \sum_{i=1}^{n} H\left(Y_i\right) - \sum_{i=1}^{n} H\left(Y_i \mid X_i\right) \\
&\le \sum_{i=1}^{n} I\left(X_i;Y_i\right) \le nC
\end{aligned}
$$

Part C: Converse to the Channel Coding Theorem

Converse to the channel coding theorem

- No rate $R > C$ is achievable.

- Suppose a $(2^{nR}, n)$ code and a uniform distribution on $\mathcal{W}$.

Proof

$$nR = H(W) = H(W \mid \hat{W}) + \underbrace{H(W) - H(W \mid \hat{W})}$$

$$= H(W \mid \hat{W}) + I(W;\hat{W})$$

$$\le 1 + P(\hat{W} \ne W)nR + \underbrace{I(W;\hat{W})}$$ FANO

$$\boxed{W \to X^n \to Y^n \to \hat{W} \Rightarrow I\left(W;\hat{W}\right) \le I\left(X^n;Y^n\right)}$$

$$\le 1 + P(\hat{W} \ne W)nR + I(X^n;Y^n)$$

$$\le 1 + P(\hat{W} \ne W)nR + nC$$

Conclusion

$$\boxed{nR \le 1 + P(\hat{W} \ne W)nR + nC}$$

– dividing by n $\quad R \le \frac{1}{n} + P(\hat{W} \ne W)R + C$

– Dividing by R and rearranging

$$P(\hat{W} \ne W) \ge 1 - \frac{C}{R} - \frac{1}{nR}$$

– Hence, for large n, if $R > C,\ P(\hat{W} \ne W) > 0$ strictly. Then it must be true for all n.

$P(\hat{W} \ne W) =$ average of λ_i's $\qquad$ so $\lambda^{(n)} = \max \lambda_i > 0$ as well

Converse to the channel coding theorem

- No rate $R > C$ is achievable.