# LECTURE 13

## Last time:

- Strong coding theorem
- Revisiting channel and codes
- Bound on probability of error
- Error exponent

## Lecture outline

- Fano's Lemma revisited
- Fano's inequality for codewords
- Converse to the coding theorem

Reading: Sct. 8.9.

# Fano's lemma

Suppose we have r.v.s $X$ and $Y$, Fano's lemma bounds the error we expect when estimating $X$ from $Y$

We generate an estimator of $X$ that is $\widehat{X} = g(Y)$.

Probability of error $P_e = Pr(\widehat{X} \neq X)$

Indicator function for error $\mathbf{E}$ which is 0 when $X = \widehat{X}$ and 1 otherwise. Thus, $P_e = P(\mathbf{E} = 1)$

Fano's lemma:

$$H(\mathbf{E}) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$$

We now need to consider the case where we are dealing with codewords

Want to show that vanishingly small probability of error is not possible if the rate is above capacity

## Fano's inequality for code words

An error occurs when the decoder makes the wrong decision in selecting the message that was transmitted

Let $M \in \{1, 2, \ldots, 2^{nR}\}$ be the transmitted message and let $\widehat{M}$ be the estimate of the received message from $\underline{Y}^n$

$M$ is uniformly distributed in $\{1, 2, \ldots, 2^{nR}\}$ and consecutive message transmissions are IID (thus, we do not make use of a number of messages, but consider a single message transmission)

The probability of error for a codebook for transmission of $M$is $P_{e,M} = P(M \neq \widehat{M}) = E_{\underline{Y}^n}[P(M \neq \widehat{M}|\underline{Y}^n)]$

Consider an indicator variable $\mathbf{E} = 1$ when an error occurs and $\mathbf{E} = 0$ otherwise

## Fano's inequality for code words

$$
\begin{aligned}
& H(\mathbf{E}, M|\underline{Y}) \\
= \ & H(M|\underline{Y}) + H(\mathbf{E}|M, \underline{Y}) \\
= \ & H(M|\underline{Y}) \\
= \ & H(\mathbf{E}|\underline{Y}) + H(M|\mathbf{E}, \underline{Y}) \\
\leq \ & 1 + H(M|\mathbf{E}, \underline{Y})
\end{aligned}
$$

Let us consider upper bounding the RHS

$H(M|\mathbf{E}, \underline{Y})$

we are not averaging over codebooks
as for the coding theorem,
but are considering a specific codebook

$$
\begin{aligned}
= \ & H(\underline{X}|\mathbf{E}, \underline{Y}) \\
= \ & E_{M,\underline{Y}}[P(M \neq \widehat{M}|\underline{Y})]H(\underline{X}|\mathbf{E} = 1, \underline{Y}) \\
+ \ & (1 - E_{M,\underline{Y}}[P(M \neq \widehat{M}|\underline{Y})]) \\
& H(\underline{X}|\mathbf{E} = 0, \underline{Y}) \\
= \ & P_e H(\underline{X}|\mathbf{E} = 1, \underline{Y}) \\
\leq \ & P_e H(\underline{X}|\mathbf{E} = 1) \\
\leq \ & P_e \log(|\mathcal{M}| - 1)
\end{aligned}
$$

## Fano's inequality for code words

Given the definition of rate, $|\mathcal{M}| = 2^{nR}$, so

$H(M|\mathbf{E}, \underline{Y}) \leq P_e nR + 1$

Hence

$$
\begin{aligned}
& H(M|\underline{Y}) \\
\leq \; & P_e nR
\end{aligned}
$$

For a given codebook, $M$ determines $\underline{X}$, so

$H(\underline{X}|\underline{Y}) = H(M|\underline{Y}) \leq 1 + P_e nR$

for a DMC with a given codebook and uniformly distributed input messages

## From Fano's inequality for code words to the coding theorem converse

We now want to relate this to mutual information and to capacity

Strategy:

- will need to have mutual information expressed as $H(M) - H(M|\underline{Y})$

- rate will need to come in play - try the fact that $H(M) = nR$ for uniformly distributed messages

- will need capacity to come into play. We remember that combining the chain rule for entropies and the fact that conditioning reduces entropy yields the fact that for a DMC $I(\underline{X}^n; \underline{Y}^n) \leq nC$

# Converse to the channel coding theorem

Consider some sequence of codebooks $(2^{nR}, n)$, indexed by $n$, such that the maximum probability of error over each codebook goes to 0 as $n$ goes to $\infty$

Assume (we'll revisit this later) that the message $M$ is drawn with uniform PMF from $\{1, 2, \ldots, 2^{nR}\}$

Then $nR = H(M)$

Also

$$
\begin{aligned}
H(M) &= H(M|\underline{Y}) + I(M; \underline{Y}) \\
&= H(M|\underline{Y}) + H(\underline{Y}) - H(\underline{Y}|M) \\
&= H(M|\underline{Y}) + H(\underline{Y}) - H(\underline{Y}|\underline{X}) \\
&= H(M|\underline{Y}) + I(\underline{X}; \underline{Y}) \\
&\leq 1 + P_e nR + nC
\end{aligned}
$$

Hence $R \leq \frac{1}{n} + P_e R + C$

## Converse to the channel coding theorem

Letting $n$ go to $\infty$, we obtain that $R \leq C$ (since the maximum probability of error goes to 0 by our assumption)

Moreover, we obtain the following bound on error: $P_e \geq 1 - \frac{C}{R} - \frac{1}{nR}$

Note:

- for $R < C$, the bound has a negative RHS, so does not bound probability of error in a way that is inconsistent with forward coding theorem

- for $R > C$, bound becomes $1 - \frac{C}{R}$ for large $n$, but $1 - \frac{C}{R} - \frac{1}{R}$ is always lower bound

- as $R$ goes to infinity, bound becomes 1, so is tight bound

- RHS of bound *does not* vary with $n$ in the way we would expect, since the bound increases with $n$

## Revisiting the message distribution

We have assumed that we can select the messages to be uniformly distributed

This is crucial to get $H(M) = nR$

Does the converse only work when the messages are uniformly distributed?

Let us revisit the consequences of the AEP

# Consequences of the AEP: the typical set

Definition: $A_\epsilon^{(n)}$ is a typical set with respect to $P_X(x)$ if it is the set of sequences in the set of all possible sequences $\underline{x}^n \in \underline{\mathcal{X}}^n$ with probability:

$$2^{-n(H(X)+\epsilon)} \leq P_{\underline{X}^n}(\underline{x}^n) \leq 2^{-n(H(X)-\epsilon)}$$

equivalently

$$H(X) - \epsilon \leq -\frac{1}{n}\log(P_{\underline{X}^n}(\underline{x}^n)) \leq H(X) + \epsilon$$

We shall use the typical set to describe a set with characteristics that belong to the majority of elements in that set.

## Consequences of the AEP: the typical set

Why is it typical? The probability of being more than $\delta$ away from $H(X)$ goes can be arbitrarily close to 0 as $n \rightarrow \infty$, hence

$$Pr(A_{\epsilon}^{(n)}) \geq 1 - \epsilon$$

We can select $\epsilon$ to be arbitrarily small, so that the distribution of messages is arbitrarily close to uniform in the typical set

The max of the probability of error must be bounded away from 0 in the typical set for the max of the probability of error to be bounded away from 0

The probability of error is dominated by the probability of the typical set as we let $\epsilon > 0$

6.441 Information Theory
Spring 2010