Fano's Inequality

Suppose you see Y, the output of some noisy channel, and you want to guess what the input to the channel must have been. Let your guess $\hat{X}$ be some function of your observation of Y, that is, $\hat{X} = g(Y)$. Note that $X \to Y \to \hat{X}$ forms a Markov chain.

Fano's inequality relates the probability that your guess is wrong $P[\hat{X} \neq X])$ to $H(X|Y)$: the uncertainty you have about the channel's input X when you are only given the output Y.

Theorem: Fano's Inequality

Let P_{XY} an arbitrary joint distribution of random variables X and Y, and let $\hat{X} = g(Y)$ for some function g. Furthermore, define $p_e := P[\hat{X} \neq X]$ to be the probability of error. Then

$$H(X|Y) \leq p_e \cdot \log(|\mathcal{X}| - 1) + h(p_e).$$

Since we know that $0 \leq p_e \leq 1$, and thus $h(p_e) \leq 1$ we may rewrite Fano's inequality as

$$p_e \geq \frac{H(X|Y) - 1}{\log(|\mathcal{X}| - 1)}.$$

Proof

Define the random variable E to be 0 whenever $\hat{X} = X$, and 1 otherwise. (In other words, E indicates whether an error has occurred in guessing the input.)

Observe the following relations between E, X, and $\hat{X}$:

1. $H(E|X\hat{X}) = 0$ (since E is a function of X and $\hat{X}$).
2. $H(E|\hat{X}) \leq H(E) = h(p_e)$ (by general properties of conditional entropy).
3. $H(X|\hat{X}, E = 0) = 0$ (if you know that the guess was correct, you can infer the original input from the guess).
4. $H(X|\hat{X}, E = 1) \leq \log(|\mathcal{X}| - 1)$ (if you know that the guess was incorrect, you only know that the correct input was one of the $|\mathcal{X}| - 1$ other options).

These observations allow us to derive the inequality:

$$\begin{aligned}
H(X|Y) &\leq H(X|\hat{X}) && \text{(by the data-process} \\
&= H(E|\hat{X}) + H(X|E\hat{X}) && \text{(by entropy diagrams and ol} \\
&\leq h(p_e) + H(X|E\hat{X}) && \text{(by ol} \\
&= h(p_e) + P_E(0) \cdot H(X|\hat{X}, E = 0) + P_E(1) \cdot H(X|\hat{X}, E = 1) \\
&= h(p_e) + 0 + P_E(1) \cdot H(X|\hat{X}, E = 1) && \text{(by ol} \\
&\leq h(p_e) + p_e \cdot \log(|\mathcal{X}| - 1) && \text{(by ob}
\end{aligned}$$