Introduction to Information Theory, Fall 2019

Practice problem set #8

You do **not** have to hand in these exercises, they are for your practice only.

1. **Fano's inequality** In the lecture Fano's inequality was stated: if $S \to Y \to \hat{S}$ is a Markov chain and S takes values on the alphabet $\mathcal{A}$ and we denote by $p_e = \Pr(S \neq \hat{S})$ then

$$H(\{p_e, 1-p_e\}) + p_e \log(|\mathcal{A}|) \geqslant H(S|Y).$$

In this exercise we will go through the proof.

(a) Define the random variable E by $E = 0$ if $S = \hat{S}$ and $E = 1$ if $S \neq \hat{S}$. Use the chain rule to show that

$$H(E, S|\hat{S}) = H(S|\hat{S})$$

and

$$\begin{aligned} H(E, S|\hat{S}) &= H(E|\hat{S}) + H(S|E, \hat{S}) \\ &\leqslant H(\{p_e, 1-p_e\}) + p_e \log(|\mathcal{A}|). \end{aligned}$$

Hint: $H(E|S, \hat{S}) = 0$ *and* $H(S|\hat{S}, E = 0) = 0$*, do you see why?*

(b) Use this to prove Fano's inequality.
Hint: use the data processing inequality.

(c) Adapt the proof to show that if S and $\hat{S}$ *both* take values on the same alphabet $\mathcal{A}$ then

$$H(\{p_e, 1-p_e\}) + p_e \log(|\mathcal{A}| - 1) \geqslant H(S|Y).$$

Hint: in the proof, given $\hat{S}$ *and* E*, how many values can* S *take?*

2. **Converse to the noisy coding theorem** We now have all the ingredients for proving the converse statement in the noisy coding theorem. As in the lecture we let

$$S \to X^N \to Y^N \to \hat{S}$$

be a channel with a coding and decoding, with rate $R = K/N$. We denote the capacity of the channel by C.

(a) Show that

$$H(S|Y^N) \geqslant K - NC.$$

Hint: use the data processing inequality and the fact (which you will prove in the homework) that $I(X^N : Y^N) \leqslant NC$.

(b) On the other hand, show that

$$H(S|Y^N) \leqslant 1 + p_B K$$

where p_B the probability of block error.
Hint: use Fano's inequality.

(c) Conclude that

$$p_B \geqslant 1 - \frac{C}{R} - \frac{1}{NR}$$

so if the rate is larger than the capacity, p_B is lower bounded by a positive constant as N goes to infinity.

3. **Rate distortion and bit error** Now we assume we are in the same setting of a channel with an encoder and decoder

$$S^K \to X^N \to Y^N \to \hat{S}^K$$

but we also assume that the number of messages is 2^K with K an integer, and we think of $S^K = (S_1 \dots S_K)$ and $\hat{S}^K = (\hat{S}_1 \dots \hat{S}_K)$ as bitstrings of length K . The goal of this problem is to bound what rates are possible if we allow a finite *bit error probability*. We define the probability of bit error as

$$p_b = \frac{1}{K}\sum_{i=1}^{K} \Pr(\hat{S}_i \neq S_i)$$

(which assumes a uniform distribution over the messages).

(a) Argue that $p_b \leqslant p_B$.

(b) Show that

$$H(\{p_b, 1-p_b\}) \geqslant \frac{1}{K}\sum_{i=1}^{K} H(\{p_{b,i}, 1-p_{b,i}\})$$

where $p_{b,i} = \Pr(\hat{S}_i \neq S_i)$

(c) Show that

$$H(\{p_b, 1-p_b\}) \geqslant \frac{1}{K}\sum_{i=1}^{K} H(S_i|Y^N) \geqslant \frac{1}{K}H(S^K|Y^N).$$

Hint: use the version of Fano's inequality in 1(c).

(d) Show that

$$H(\{p_b, 1-p_b\}) \geqslant \frac{1}{K}(H(S) - I(X^N : Y^N)) \geqslant 1 - \frac{C}{R}$$

and use this to conclude that

$$R \leqslant \frac{C}{1 - H(\{p_b, 1-p_b\})}.$$

Hint: use the data processing inequality and use that $I(X^N : Y^N) \leqslant \sum_{i=1}^{N} I(X_i : Y_i)$.