ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 36 — Information Theory and Coding

Final exam solutions — Feb. 8, 2022

Problem 1.

(a) $I_1 = \sum_{x,y} p_1(x)W(y|x)\log\left(\frac{p_1(x)W(y|x)}{p_1(x)q_1(y)}\right)$
$= \sum_x p_1(x)\sum_y W(y|x)\log\left(\frac{W(y|x)}{q_1(y)}\right) = \sum_x p_1(x)D\big(W(\cdot|x)\;\|\;q_1\big)$.
Also note that,

$$
\begin{aligned}
I_1 &= \sum_x p_1(x)\sum_y W(y|x)\log\left(\frac{W(y|x)}{q_1(y)}\right) &(1)\\
&= \sum_x p_1(x)\sum_y W(y|x)\log\left(\frac{W(y|x)}{q_2(y)}\frac{q_2(y)}{q_1(y)}\right) &(2)\\
&= \sum_x p_1(x)\sum_y W(y|x)\log\left(\frac{W(y|x)}{q_2(y)}\right) - \sum_y q_1(x)\log\left(\frac{q_1(y)}{q_2(y)}\right) &(3)\\
&= \sum_x p_1(x)D\big(W(\cdot|x)\;\|\;q_2\big) - D(q_1\|q_2) &(4)
\end{aligned}
$$

(b) By KKT conditions, since p_2 is a capacity achieving distribution,

$$
\begin{cases} D(W(\cdot|x)\|q_2) = C(W), & \text{if } p_2(x) > 0 \\ D(W(\cdot|x)\|q_2) \le C(W), & \text{if } p_2(x) = 0 \end{cases}
$$

Therefore, $\sum_x p_1(x)D(W(\cdot)\|q_2) \le C(W)$.

(c) Using part (a) and part (b), $I_1 = \sum_x p_1(x)D\big(W(\cdot|x)\;\|\;q_2\big) - D(q_1\|q_2)$
$\le \sum_x p_1(x)D\big(W(\cdot|x)\;\|\;q_2\big) \le C(W)$. If p_1 is capacity achieeving $I_1 = C(w)$ which means we need equality in the first inequality above. This implies $D(q_1\|q_2) = 0 \implies q_1 = q_2$.

(d) $I(X;Y) = H(Y) - H(Y|X) = H(Y) \le 1$. ($H(Y|X) = 0$ since the channel is deterministic.) Any input distribution which imposes a uniform distribution in the output achieves a capacity. Therefore, any input distribution which has $p(0) = 1/2$ achieve the capacity.

PROBLEM 2.

(a) The condition α suggests will be true with very low probability, because if W_1 is used $(\text{Enc}(m), y^n) \in T_2$ with exponentially decreasing probability ($\propto 2^{-nD(p^{(2)}_{XY})\|p^{(1)}_{XY})}$)and therefore, $\Pr\left((\text{Enc}(m), y^n) \in T_1 \cap T_2\right)$ is even lower. Therefore, even the true message will be decoded with low probability. As for β the true message will be typical with high probability. We know that if W_1 is used, an unsent message-output pair will be typical with respect to T_1 with low probability. Hoping that it has no specific reason to look typical as if it had been sent from W_2 the unsent message will not be decoded.

(b) Assume that the message is chosen uniformly and indepently from the codebook. Then,

$$\Pr(\text{Error}) = \frac{1}{M} \sum_{m=1}^{M} \Pr(\text{Error} \mid M = m) \tag{5}$$

Due to symmetry,

$$= \Pr(\text{Error} \mid M = 1) \tag{6}$$

$$= \Pr((\text{Enc}(1), Y^n) \notin T_1 \cup T_2 \textbf{ or } \exists m \neq 1 (\text{Enc}(m), Y^n) \in T_1 \cup T_2) \tag{7}$$

$$\leq \Pr((\text{Enc}(1), Y^n) \notin T_1 \cup T_2) + \sum_{m \neq 1} \Pr((\text{Enc}(m), Y^n) \in T_1 \cup T_2) \tag{8}$$

$$\leq \Pr((\text{Enc}(1), Y^n) \notin T_1) + \sum_{m \neq 1} \Pr((\text{Enc}(m), Y^n) \in T_1) + \Pr((\text{Enc}(m), Y^n) \in T_2) \tag{9}$$

Note that for any $m \neq 1$, $\text{Enc}(1), Y^n, \text{Enc}(m)$ has the same distribution as $(X^n, Y^n, \tilde{X}^n)$

$$= \Pr(X^n, Y^n \notin T_1) + (M-1)\Pr(\tilde{X}^n, Y^n \in T_1) + (M-1)\Pr(\tilde{X}^n, Y^n \in T_2) \tag{10}$$

(c)

$$D(p_2 \| q_1) = \mathbb{E}_{p_X W_2}\left[\log\left(\frac{p_X W_2}{p_X p_Y^{(1)}(y)}\right)\right] \tag{11}$$

$$= \mathbb{E}_{p_X W_2}\left[\log\left(\frac{p_X W_2}{p_X p_Y^{(2)}(y)}\right)\right] + D(p_Y^{(2)} \| p_Y^{(1)}) \tag{12}$$

$$\geq \mathbb{E}_{p_X W_2}\left[\log\left(\frac{p_X W_2}{p_X p_Y^{(2)}(y)}\right)\right] \tag{13}$$

$$= I_2 \tag{14}$$

(d) Let us construct a random codebook $\mathcal{C}$ using the probability distribution p_X. It is enough to show that,

$$\mathbb{E}[\max(p_{e,1}(\mathcal{C}), p_{e,2}(\mathcal{C})] \to 0$$

as $n \to \infty$, where the randomness is over the random codebook and $p_{e,i}(\mathcal{C})$ denotes error probability under channel i given codebook $\mathcal{C}$. Note that,

$$\mathbb{E}[\max(p_{e,1}(\mathcal{C}), p_{e,2}(\mathcal{C}))] \leq \mathbb{E}[p_{e,1}(\mathcal{C})]] + \mathbb{E}[p_{e,2}(\mathcal{C})]]$$

If we show that both of the terms in the right hand side go to 0 we are done. Let us do the first term. The proof for the second term is similar. Note that the term $\mathbb{E}[p_{e,1}(\mathcal{C})]]$ is the probability of error when $\mathcal{C}$ is chosen randomly and the first channel is used. We want to show this probability of error goes to 0.
In general if Z^n is i.i.d. samples from q the probability that it looks p typical is $\sim 2^{-nD(p\|q)}$.
The first term in Eq. (10) goes to 0. The second term scales as $M2^{-nD(p_{XY}^{(1)}\|p_X p_Y^{(1)})} = M2^{-nI_1}$ and the third term scales as $M2^{-nD(p_{XY}^{(2)}\|p_X p_Y^{(1)})} = M2^{-nD(p_2\|q_1)} \leq M2^{-nI_2}$. Provided that $M \leq 2^{n\min(I_1,I_2)}$, all of the terms go to 0. Since encoder is chosen randomly, for any $R < \min(I_1, I_2)$ there must exists a decoder with error probability arbitrarily close to 0 and with rate at least R.

Problem 3.

(a)

$$H(N) - H(G) = \sum_k p_N(k) \log \frac{1}{p_N(k)} - \sum_k p_G(k) \log \frac{1}{p_G(k)} \tag{15}$$

Note that $\log p_N(k)$ is a first order polynomial in k. Since $\mathbb{E}[N] = \mathbb{E}[G]$,

$$\begin{aligned} &= \sum_k p_G(k) \log \frac{1}{p_N(k)} - \sum_k p_G(k) \log \frac{1}{p_G(k)} && (16) \\ &= D(p_G \| p_N). && (17) \end{aligned}$$

(b) Given G identify the random variable N with parameter q such that $\mathbb{E}[N] = \mathbb{E}[G]$. Then, $H(G) = H(N) - D(p_G \| p_N) \leq H(N) = f(\mathbb{E}[N]) = f(\mathbb{E}[G])$.

(c) We claim that the optimal questioning strategy is $S = (1, 2, ..., K)$. We will prove this by contradiction. Suppose that the optimal questioning strategy S is different. Then, $\exists i, j$ such that $p_i > p_j$ but "$U = i$?" is asked before "$U = j$?". That is $S = (..i, ..., j, ..)$ for $i > j$. Consider the strategy S' where the position of i and j are switched and everything else remains the same. With probability p_i S' will take δ more guesses compared to S. However, with probability p_j it will take δ less guesses compared to S. In average, S' will be shorter because $(p_i - p_j)\delta > 0$.

(d) WLOG, we can still assume $p_1 \geq p_2 \cdots \geq p_K$. Otherwise reorder the elements. For every k, $\Pr(G \geq k) \geq p_{k+1} + p_{k+2} + \cdots p_K$, because whatever the guessing strategy is, $G > k$ if and only if one of the last $K - k$ guesses is true. The last $K - k$ guesses cannot have a probability lower than $p_{k+1} + p_{k+2} + \cdots p_K$. The strategy we found in (c) satisfy, for all k, the bounds with equality.

(c,d) (alternative proof)

$$\mathbb{E}[G] = \sum_{k=1}^{K} p_G(k) k = \sum_{k=1}^{K} \Pr(G \geq k) \tag{18}$$

Therefore, $\min \mathbb{E}[G] = \min \sum_{k=1}^{K} \Pr(G \geq k) \geq \sum_{k=1}^{K} \min \Pr(G \geq k)$. For a fixed k the strategy which minimizes $\Pr(G \geq k)$ leaves the least probable $K - k$ letters to be guessed after the others. The guessing strategy which guesses in the order $1, 2, ..., K$ achieves the lower bound for every k. Therefore, it minimizes $\mathbb{E}[G]$.

(e) Note that for any fixed guessing strategy G is just a permutation of U therefore, $H(G) = H(U)$.

(f) By part (b), $f(\mathbb{E}[G_n | U^{n-1} = u^{(n-1)}]) \geq H(p_{G_n | U^{n-1} = u^{n-1}}) = H(p_{U_n | U^{n-1} = u^{n-1}}) = H(U_n | U^{n-1} = u^{n-1})$. Now take average over U^{n-1}, $H(U_n | U^{n-1}) \leq \mathbb{E}[f(\mathbb{E}[G_n | U^{n-1}])] \leq f(\mathbb{E}[\mathbb{E}[G_n | U^{n-1}]]) = f(\mathbb{E}[G_n])$. Conclusion follows from the fact that for every n, $H \leq H(U_n | U^{n-1})$ for stationary processes.

Problem 4.

(a) Let $\mathcal{C}$ be a binary code. Suppose we have a bound $\text{bound}_0(n, d)$ which applies to all binary codes with minimum distance at least d.

Suppose $x = (x_1, x_2, ..., x_n)$. We define $f_{n'}(x) = (x_1, x_2, ...x_{n-n'})$ and $f_{n'}(\mathcal{C}) = \{f_{n'}(x) : x \in \mathcal{C}\}$. Fix some $y \in f_{n'}(\mathcal{C})$ and look at the inverse image. These are the codewords that have the same $n - n'$ bit prefix and they are at least d apart from each other. Since the first $n - n'$ bits are the same the remaining n' bits have to be at least d apart as well. Therefore, the number of elements in the inverse image of y have to be less than or equal to $\text{bound}_0(n', d)$. This bound is true for every y. Since there are at most $2^{n-n'}$ elements in $f_{n'}(\mathcal{C})$, $M \leq 2^{n-n'}\,\text{bound}_0(n', d)$, for every $0 \leq n' \leq n$.

(b) Suppose $M \geq 2$ then their distance have to be at most n. Therefore, if $d > n$ we can have at most 1 element in the code. For $d \leq n$ infinity is also an upper bound because it is bigger than any natural number.

(c) Choose $n' = d - 1$. For this, n', $\text{bound}_0(n', d) = 1$. Therefore, $\text{bound}_1(n, d) \leq 2^{n-d+1}$. This is actually the minimum because $2^{n-n'}$ increases as one further decreases n'. This is the same as the Singleton bound.

(d)

$$\sum_{m=1}^{M}\sum_{m'=1}^{M} \mathbb{1}\{c_{m_i} \neq c_{m'i}\} = \sum_{m:c_{m_i}=1}\sum_{m'=1}^{M} \mathbb{1}\{c_{m_i} \neq c_{m'i}\} + \sum_{m:c_{m_i}=0}\sum_{m'=1}^{M} \mathbb{1}\{c_{m_i} \neq c_{m'i}\} \tag{19}$$

$$= \sum_{m:c_{m_i}=1}\sum_{m'=1}^{M} \mathbb{1}\{0 = c_{m'i}\} + \sum_{m:c_{m_i}=0}\sum_{m'=1}^{M} \mathbb{1}\{1 = c_{m'i}\} \tag{20}$$

$$= M_i(M - M_i) + (M - M_i)M_i \tag{21}$$

$$= 2M_i(M - M_i) \tag{22}$$

maximize the expression over M_1 to get,

$$\leq \frac{M^2}{2}. \tag{23}$$

(e)

$$n\frac{M^2}{2} \geq \sum_{i=1}^{n}\sum_{m=1}^{M}\sum_{m'=1}^{M} \mathbb{1}\{c_{m_i} \neq c_{m'i}\} \tag{24}$$

$$= \sum_{m=1}^{M}\sum_{m'=1}^{M}\sum_{i=1}^{n} \mathbb{1}\{c_{m_i} \neq c_{m'i}\} \tag{25}$$

$$= \sum_{m=1}^{M}\sum_{m'=1}^{M} d_H(c_m, c_{m'}) \tag{26}$$

$$\geq \sum_{(m,m'):m\neq m'} d_{\min}(\mathcal{C}) \tag{27}$$

$$= M(M-1)d. \tag{28}$$

(f) Note that $2d = n$. The bound says $14 = 2d \leq \frac{nM}{M-1} = 14\frac{M}{M-1}$. That is, $1 \leq M/(M-1)$. From this M is not bounded. Let us improve the bound via (a). The method works because the bound is non-increasing in d. Choosing $n' = 13$ minimizes the bound. It gives,$M \leq 2^{14-13}\frac{14}{14-13} = 28$. The Singleton Bound gives, $M \leq 2^{n-d+1} = 2^{14-7+1} = 256$. Sphere Packing Bound gives, $M \leq \lfloor 2^{14}/V_3 \rfloor = 34$. Therefore, the new bound strictly improves Singleton and Sphere packing bounds.