ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 36 Information Theory and Coding

Final exam solutions Jan. 30, 2020

PROBLEM 1.

a)

$$\begin{aligned} H(AB) + H(BC) &= H(B) + H(A|B) + H(BC) \\ &\geq H(B) + H(A|BC) + H(BC) \\ &= H(B) + H(ABC). \end{aligned}$$

b) Choose $B = X_{\mathcal{S}\cap\mathcal{T}}$, $A = X_{\mathcal{S}\setminus\mathcal{T}}$, $C = X_{\mathcal{T}\setminus\mathcal{S}}$. Then

$$H(AB) = H(X_{\mathcal{S}}), \quad H(BC) = H(X_{\mathcal{T}}), \quad H(ABC) = H(X_{\mathcal{S}\cup\mathcal{T}}).$$

Direct application of (a) yields the result.

c) In the hint, it is given that the left-hand side is the average of $H(X_{i_{k+1}}|X_{i_2}, \dots, X_{i_k})$ over all permutations $(i_1, \dots, i_n)$ of $(1, \dots, n)$. Observe

$$H(X_{i_{k+1}}|X_{i_2}, \dots, X_{i_k}) = H(X_{i_2}, \dots, X_{i_k}, X_{i_{k+1}}) - H(X_{i_2}, \dots, X_{i_k})$$

and

$$\frac{1}{n!} \sum_{(i_1,\dots,i_n)\in\pi(1,\dots,n)} H(X_{i_{k+1}}|X_{i_2}, \dots, X_{i_k})$$

$$= \frac{1}{n!} \sum_{(i_1,\dots,i_n)\in\pi(1,\dots,n)} H(X_{i_2}, \dots, X_{i_k}, X_{i_{k+1}}) - \frac{1}{n!} \sum_{(i_1,\dots,i_n)\in\pi(1,\dots,n)} H(X_{i_2}, \dots, X_{i_k}).$$

Consider the first sum $\frac{1}{n!}\sum H(X_{i_2}, \dots, X_{i_k}, X_{i_{k+1}})$. Note that for any set $\mathcal{S} \subset \{1, \dots, n\}$ with $|\mathcal{S}| = k$, $H(X_{\mathcal{S}})$ is counted $(n-k)!k!$ times in the above sum. Therefore,

$$\frac{1}{n!} \sum_{(i_1,\dots,i_n)\in\pi(1,\dots,n)} H(X_{i_2}, \dots, X_{i_k}, X_{i_{k+1}}) = \frac{1}{\binom{n}{k}} \sum_{\mathcal{S}:|\mathcal{S}|=k} H(X_{\mathcal{S}}) = H_k.$$

Likewise, we have $\frac{1}{n!}\sum H(X_{i_2}, \dots, X_{i_k}) = H_{k-1}$ for the second sum. With a similar reasoning, we see that the right-hand side is the average of $H(X_{i_{k+1}}|X_{i_1}, \dots, X_{i_k})$ over all permutations $(i_1, \dots, i_n)$ of $(1, \dots, n)$.

Since

$$H(X_{i_{k+1}}|X_{i_2}, \dots, X_{i_k}) \geq H(X_{i_{k+1}}|X_{i_1}, \dots, X_{i_k}),$$

we obtain

$$\frac{1}{n!} \sum_{(i_1,\dots,i_n)\in\pi(1,\dots,n)} H(X_{i_{k+1}}|X_{i_2}, \dots, X_{i_k}) \geq \frac{1}{n!} \sum_{(i_1,\dots,i_n)\in\pi(1,\dots,n)} H(X_{i_{k+1}}|X_{i_1}, \dots, X_{i_k})$$

which is equivalent to

$$H_k - H_{k-1} \geq H_{k+1} - H_k.$$

d) Let $a_k := H_k - H_{k-1}$. Using the hint, we obtain

$$\frac{H_k}{k} = \frac{1}{k}\sum_{i=1}^{k} a_i, \quad \frac{H_{k+1}}{k+1} = \frac{1}{k+1}\sum_{i=1}^{k+1} a_i$$

i.e., averages of $(a_i)_{i=1}^{k}$ and $(a_i)_{i=1}^{k+1}$ respectively. From part (c) we know that the sequence (a_k) is non-increasing, which implies $a_{k+1} \leq a_i$ for all $i = 1, \ldots, k$. It is known the average of the sequence $(a_i)_{i=1}^{k+1}$ is smaller than the average of $(a_i)_{i=1}^{k}$ if a_{k+1} is smaller than every other term in the sequence. This proves the statement.

If the above fact is not obvious, one can proceed with

$$\begin{aligned}
\frac{H_k}{k} - \frac{H_{k+1}}{k+1} &= \frac{1}{k}\sum_{i=1}^{k} a_i - \frac{1}{k+1}\sum_{i=1}^{k+1} a_i \\
&= \frac{\sum_{i=1}^{k} a_i - k a_{k+1}}{k(k+1)} \\
&= \frac{\sum_{i=1}^{k} (a_i - a_{k+1})}{k(k+1)} \geq 0.
\end{aligned}$$

PROBLEM 2.

(a) Observe that $H(Z_1^n|W) \le H(Z_1^n) \le H(Z_1^n, W)$. We also have

$$\begin{aligned} H(Z_1^n|W) &= \frac{1}{3}\big(H(Z_1^n|W=0) + H(Z_1^n|W=1) + H(Z_1^n|W=2)\big) \\ &= \frac{1}{3}H(Z_1^n|W=2) = \frac{n}{3}. \end{aligned}$$

and

$$\begin{aligned} H(Z_1^n, W) &= H(Z_1^n|W) + H(W) \\ &= \frac{n}{3} + \log 3. \end{aligned}$$

Taking the limit for both upper and lower bounds, we obtain

$$\begin{aligned} \lim_n \frac{1}{n}H(Z_1^n|W) \le \lim_n \frac{1}{n}H(Z_1^n) &\le \lim_n \frac{1}{n}H(Z_1^n, W) \\ \frac{1}{3} \le \lim_n \frac{1}{n}H(Z_1^n) &\le \frac{1}{3}. \end{aligned}$$

Therefore, $\lim_n \frac{1}{n}H(Z_1^n) = \frac{1}{3}$.

(b) $I(X^n; Y^n) = H(Y^n) - H(Y^n|X^n) = H(Y^n) - H(Z^n)$. Note that $H(Y^n) \le n$ and equality holds if and only if Y_is are independently and uniformly distributed. This is attained when X_is are also independently and uniformly distributed. We now verify this claim.

- If $W = 0$ or $W = 1$, the noise Z_1^n is fixed and $Y_1^n = X_1^n + 0^n$ or $Y_1^n = X_1^n + 1^n$. One can see that Y_is are independently and uniformly distributed if X_is are also independently and uniformly distributed.
- If $W = 2$, then Z_is are i.i.d. and the output Y_1^n will be independently and uniformly distributed and will also be independent of the input X_1^n.

Therefore $p_X(X_1^n = x_1^n) = \frac{1}{2^n}$, for all $x_1^n = \{0,1\}^n$ maximizes $I(X^n; Y^n)$. In this case,

$$C_n = 1 - H(Z^n)/n.$$

(c) Using part (a), we have $\lim_n C_n = 1 - \lim_n H(Z^n)/n = \frac{2}{3}$.

(d) Suppose we have two codewords as we want to send one bit of information. When $W = 2$, the output is independent of the input. Therefore, the receiver cannot do better than choosing one of the codewords randomly, which implies that the error probability is $\frac{1}{2}$. Since $W = 2$ with probability $\frac{1}{3}$, we see that the error probability for any code is greater than $\frac{1}{6}$.

(e) The capacity is zero as the error probability cannot be made arbitrarily small.

Problem 3.

a) Consider any code $\mathcal{C}$ with $|\mathcal{C}| = 2^{nR}$ and error probability p_e. Taking the hint, we will need to show that :

   1) There is a k such that $|\mathcal{C}_k| \geq 2^{nR}/(n+1)$, which implies that $\log|C_k|/n = R' \geq R - \frac{\log(n+1)}{n}$. This is due to the fact that we have 2^{nR} codewords and $(n+1)$ possible value of k, i.e., $k \in \{0, 1, \ldots, n\}$. Hence it is justified by the pigeonhole principle.

      You can also prove this by contradiction. If for all k we have $|\mathcal{C}_k| < 2^{nR}/(n+1)$, then $|\mathcal{C}| = \sum_k |\mathcal{C}_k| < 2^{nR}$. This contradicts the fact that $|\mathcal{C}| = 2^{nR}$.

   2) For any k, we define $\mathcal{U}_k = \{u \in \mathcal{U} : enc(u) \in \mathcal{C}_k\}$. Therefore

   $$p'_e = \max_{u \in \mathcal{U}_k} W^n(dec(Y^n) \neq u | X^n = enc(u)) \leq \max_{u \in \mathcal{U}} W^n(dec(Y^n) \neq u | X^n = enc(u)) = p_e$$

   where the inequality is because we optimize over a subset of $\mathcal{U}$.

Now, for every $R < C$, take n large enough such that $R + \log(n+1)/n < C$. As we have discussed in class, there exists a code $\mathcal{C}$ with rate $R + \log(n+1)/n$ with arbitrarily small error probability p_e. As we have proved in 1) and 2), there exists a constant-weight subset of $\mathcal{C}$, i.e. $\mathcal{C}'_k$, with rate R and smaller error probability p'_e. This implies that there exists a rate-achieving constant-weight code.

b) Consider any codewords $x^n \in \mathcal{C}$ and any possible channel output y^n. For BSC(p), we have

$$W(Y^n = y^n | X^n = x^n) = p^{\sum_{i=1}^n \mathbb{1}\{x_i \neq y_i\}} (1-p)^{\sum_{i=1}^n \mathbb{1}\{x_i = y_i\}}.$$

For $0 < p < 1$, this probability is always positive. Hence, any pair of codewords and channel output is compatible and the decoder always return ?. This implies that $C_{eo} = 0$.

For $p = 0$ or $p = 1$, for any x^n, there is only one y^n such that this probability is positive. Hence the decoder always return a correct guess and the capacity $C_{eo} = 1$.

c) As the channel is BEC, it cannot flip bits on the channel inputs. Furthermore, as we know that y^n contains j erasures and the channel is i.i.d., then the probability of this event happens is $p^j(1-p)^{n-j}$ if x^n is compatible with y^n. Hence

$$W^n(Y^n = y^n | X^n = x^n) = \begin{cases} 0 & \exists i, y_i \neq ? \text{ and } y_i \neq x_i \\ p^j(1-p)^{n-j} & \text{otherwise} \end{cases}$$

d) By Bayes' rule, we have

$$\Pr(U = u | Y^n = y^n) = \frac{W^n(Y^n = y^n | U = u) \Pr(U = u)}{\sum_{u \in U} W^n(Y^n = y^n | U = u) \Pr(U = u)} = \frac{W^n(Y^n = y^n | U = u)}{\sum_{u \in U} W^n(Y^n = y^n | U = u)}$$

where the last inequality is due to U is distributed uniformly. From c, we know that any x^n which compatible with y^n has a similar $W^n(Y^n = y^n | U = u)$ value. Therefore we have

$$\Pr(U = u | Y^n = y^n) = \frac{1}{|\{x^n \in \mathcal{C} : x^n \text{ is compatible with } y^n\}|} \leq \frac{1}{2}$$

where the last inequality is due to the fact that $T(y^n) \geq 2$.

e) Consider the following,

$$
\begin{aligned}
&= \Pr(\hat{U} \neq U) \\
&= \sum_{y^n \in B} \Pr(\hat{U} \neq U, Y^n = y^n) + \sum_{y^n : T(y^n)=1} \Pr(\hat{U} \neq U, Y^n = y^n) + \sum_{y^n : T(y^n)=0} \Pr(\hat{U} \neq U, Y^n = y^n) \\
&= \sum_{y^n \in B} \Pr(\hat{U} \neq U, Y^n = y^n)
\end{aligned}
$$

this is due to the fact $y^n : T(y^n) = 1$ is always decoded correctly and $y^n : T(y^n) = 0$ has $W^n(Y^n = y^n | X^n = x^n) = 0$ as we have shown in c). This implies

$$
\begin{aligned}
\Pr(\hat{U} \neq U) &= \sum_{y^n \in B} \big(1 - \Pr(\hat{U} = U | Y^n = y^n)\big) P(Y^n = y^n) \\
&\geq \frac{1}{2} \sum_{y^n \in B} P(Y^n = y^n) \\
&= \frac{1}{2} P(Y^n \in B)
\end{aligned}
$$

f) From e), we can deduce that $dec_{eo}(y^n) =?$ iff $y^n \in B$. Hence $P(dec_{eo}(Y^n) \neq U) \leq 2P(\hat{U} \neq U)$. This implies that $C_{eo}(W) \geq C(W)$ for BEC, because if there exists a code with vanishing p_e then there exists codes with vanishing p_{eo}.

Now, consider our expansion from e)

$$
p_e = \Pr(\hat{U} \neq U) = \sum_{y^n \in B} \big(1 - \Pr(\hat{U} = U | Y^n = y^n)\big) P(Y^n = y^n) \leq \sum_{y^n \in B} P(Y^n = y^n) = p_{eo}.
$$

In other words, if there exists a codes with vanishing p_{eo} then there exist a code with vanishing p_e. This implies that $C_{eo}(W) \leq C(W)$ for BEC.

Hence, $C_{eo}(W) = C(W)$ for BEC.

PROBLEM 4.

a) Consider codewords which achieves minimal distance $enc(a)$ and $enc(b)$, define the sets $A_{ab} = \{k : x_{i,k} = 1, x_{j,k} = 0\}$, $B_{ab} = \{k : x_{i,k} = x_{j,k} = 1\}$,and $C_{ab} = \{k : x_{i,k} = 0, x_{j,k} = 1\}$. As the code is constant-weight, we have $|A_{ab}| + |B_{ab}| = |B_{ab}| + |C_{ab}| = k$ which implies

$$d = |A_{ab}| + |C_{ab}| = 2k - 2|B_{ab}|$$

Hence, d must be an even number, as it is equal to an even number minus an even number.

A constant-weight code cannot be linear, because linear codes must contain all zero vectors with weight 0. But we define $k > 0$. Hence contradiction.

b) For any pair of distinct codewords $enc(a)$ and $enc(b)$, define A_{ab},B_{ab},C_{ab} as in a). Consider the following equality

$$\sum_{j=1}^{n} x_{a,j}x_{b,j} = |B_{ab}| = k - \frac{|A_{ab}| + |C_{ab}|}{2}.$$

As it must hold for every $a \neq b$ then

$$\sum_{j=1}^{n} x_{a,j}x_{b,j} \leq k - \min_{a^*,b^*,a^* \neq b^*} \frac{|A_{a^*b^*}| + |C_{a^*b^*}|}{2} = k - \frac{d}{2}.$$

c) This is a consequence of the Cauchy-Schwartz inequality

$$\left(\sum_{j=1}^{n} w_j 1\right) \leq \sum_{j=1}^{n} w_j^2 \sum_{j=1}^{n} 1 = n \sum_{j=1}^{n} w_j^2.$$

this implies

$$\frac{k^2M^2}{n} = \frac{1}{n}\left(\sum_{j=1}^{n} w_j 1\right) \leq \sum_{j=1}^{n} w_j^2.$$

d) We have

$$\begin{aligned}\frac{k^2M^2}{n} \leq& \sum_{j=1}^{n} \sum_{a,b \in [m]} x_{a,j}x_{b,j} \\ =& \sum_{a \neq b} \sum_{j=1}^{n} x_{a,j}x_{b,j} + \sum_{a=b} \sum_{j=1}^{n} x_{a,j}x_{b,j} \\ \leq& \sum_{a \neq b} \left(k - \frac{d}{2}\right) + \sum_{a=b} k\end{aligned}$$

where the first term is due to b) and the second term is due to its a k constant-weight code. This implies

$$\frac{k^2M^2}{n} \leq M(M-1)\left(k - \frac{d}{2}\right) + Mk$$

which is equivalent to

$$\frac{k^2M}{n} - k \leq (M-1)\left(k - \frac{d}{2}\right).$$

e) Plugging the number, we have

$$\frac{16M}{9} - 4 \leq M - 1$$

which implies

$$M \leq \frac{27}{7} = 3 + \frac{6}{7}$$

as M must be integer, then $M^* \leq 3$.

Consider the following instance of $(9, 6, 4)$ code $\{111100000, 000111100, 100000111\}$. This implies that $M^* \geq 3$.

Hence $M^* = 3$.