ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 16 Information Theory and Coding
Midterm Solutions Nov. 02, 2021

Problem 1.

(a) $I(A;B|C) = I(A;B,g(B)|C) = I(A;g(B)|C) + I(A;B|C,g(B)) \geq I(A;g(B)|C)$.

(b) $s_n = \sum_{i=1}^n \Delta_i = \sum_{i=1}^n I(Z;U_i|U^{i-1}) = I(Z;U^n) \leq H(Z)$ for all n. Since s_n is non-decreasing and bounded (by $H(Z)$), it tends to a limit s^*. This shows that $\lim_i \Delta_i = \lim_i s_i - s_{i-1} = s^* - s^* = 0$.

(c) Observe that $Z_1 \multimap Z_2 \multimap (Z_3, \dots, Z_{i+1})$. Hence

$$I(Z_1;U_2,\dots,U_{i+1}|Z_2) \leq I(Z_1;(Z_2,U_2),\dots,(Z_{i+1},U_{i+1})|Z_2)$$
$$= I(Z_1;Z_2,\dots,Z_{i+1}|Z_2) = 0.$$

This proves the statement in the Hint. Now we complete the proof by

$$I(Z_1;U_2,\dots,U_{i+1}|Z_2) = \sum_{j=1}^{i} I(Z_1;U_{j+1}|Z_2,U_2^j) = 0,$$

which shows that every term in the summation must be 0, as they must be non-negative.

(d) The first equality in part (a) implies $H(A|B,C) = H(A|B,g(B),C)$, and part (c) implies $H(U_{i+1}|Z_2,U_2^i) = H(U_{i+1}|Z_1,Z_2,U_2^i)$. Observe

$H(U_i|Z_1,U^{i-1}) \overset{(s)}{=} H(U_{i+1}|Z_2,U_2^i) \overset{(c)}{=} H(U_{i+1}|Z_1,Z_2,U_2^i) \overset{(a)}{=} H(U_{i+1}|Z_1,Z_2,U_1,U_2^i) \leq H(U_{i+1}|Z_1,U^i)$,

where (s) follows from stationarity, (c) follows from part (c), and (a) follows from part (a) with $U_1 = f(Z_1)$.

(e) First observe that $(U_1,U_2,\dots)$ is stationary, therefore its entropy rate exists and is equal to $\lim_i \frac{1}{i}H(U^i) = \lim_i H(U_i|U^{i-1}) = \mathcal{H}$. Let us write

$$a_i = H(U_i|Z_1,U^{i-1}) = H(U_i|U^{i-1}) - I(U_i;Z_1|U^{i-1}).$$

From part (b), we know $\lim_i I(U_i;Z_1|U^{i-1}) = 0$, and $\lim_i H(U_i|U^{i-1}) = \mathcal{H}$. Since both limits on the right hand side exist, the limit of a_i also exists and is equal to $\mathcal{H}$.

Since $b_i = H(U_i|U^{i-1})$ converges to the entropy rate of the process U from above, the sequence of intervals $[a_i,b_i]$ give increasingly accurate lower/upper bounds to the entropy rate, and thus we have a procedure to find the entropy rate to any desired accuracy for such processes as U. Such processes are called 'hidden Markov processes' and are good models for a large class of physical phenomena.

Problem 2.

(a) $H(U^k) = H(V_i, U_i) = H(V_i) + H(U_i|V_i)$. Therefore $kH(U^k) = \sum_{i=1}^k H(V_i) + H(U_i|V_i) \le \sum_{i=1}^k H(V_i) + H(U_i|U^{i-1}) = H(U^k) + \sum_{i=1}^k H(V_i)$.

(b) The hint suggests that for any $\mathcal{S}$ with $|\mathcal{S}| = k+1$, $\sum_{\mathcal{T}:|\mathcal{T}|=k} \mathbb{1}\{\mathcal{T} \subset \mathcal{S}\} H(U_{\mathcal{T}}) \ge kH(U_{\mathcal{S}})$. Sum both sides over $\mathcal{S}$ to obtain

$$\sum_{\mathcal{S}:|\mathcal{S}|=k+1} \sum_{\mathcal{T}:|\mathcal{T}|=k} \mathbb{1}\{\mathcal{T} \subset \mathcal{S}\} H(U_{\mathcal{T}}) \ge \sum_{\mathcal{S}:|\mathcal{S}|=k+1} kH(U_{\mathcal{S}}) = kH_{k+1}.$$

Change the order of summation on left-hand side to obtain

$$\sum_{\mathcal{S}:|\mathcal{S}|=k+1} \sum_{\mathcal{T}:|\mathcal{T}|=k} \mathbb{1}\{\mathcal{T} \subset \mathcal{S}\} H(U_{\mathcal{T}}) = \sum_{\mathcal{T}:|\mathcal{T}|=k} \sum_{\mathcal{S}:|\mathcal{S}|=k+1} \mathbb{1}\{\mathcal{T} \subset \mathcal{S}\} H(U_{\mathcal{T}})$$

$$= (n-k) \sum_{\mathcal{T}:|\mathcal{T}|=k} H(U_{\mathcal{T}}) = (n-k) H_k$$

since $\sum_{\mathcal{S}:|\mathcal{S}|=k+1} \mathbb{1}\{\mathcal{T} \subset \mathcal{S}\}$ equals to the number of subsets of size $k+1$ which contain a set $\mathcal{T}$ of size k, which is equal to $(n-k)$.

(c) Rearrange the result of part (b) to obtain $\frac{1}{k} H_k \ge \frac{1}{n-k} H_{k+1}$. Divide both sides by $\binom{n}{k}$ to obtain

$$\frac{1}{k} \frac{H_k}{\binom{n}{k}} \ge \frac{1}{n-k} \frac{H_{k+1}}{\binom{n}{k}} = \frac{1}{n-k} \frac{H_{k+1}}{\frac{n!}{(n-k)!k!}} = \frac{H_{k+1}}{\frac{n!}{(n-k-1)!k!}}$$

$$= \frac{1}{k+1} \frac{H_{k+1}}{\frac{n!}{(n-k-1)!(k+1)!}} = \frac{1}{k+1} \frac{H_{k+1}}{\binom{n}{k+1}}.$$

Problem 3.

(a) According to the notes, $l_0 = 1$, i.e., the initial dictionary contains words of length 1. Therefore, $aaaa\ldots$ will be parsed as $a, aa, aaa, \ldots$, with $w_1 = a$, $w_2 = aaa$, $w_3 = aaaa$, and so on. This shows that while w_{m+1} is being parsed, $l_m = m + 1$.

(b) Since the words added to the dictionary are 1-letter extensions of the just parsed word $l_{m+1} \leq l_m + 1$. With $l_0 = 1$, we get $l_m \leq m + 1$, i.e., the special case in (a) is the worst case.

(c) u_{n+1} is surely reconstructed upon the reception of u_{n+l_m}. We know from part (b) that $l_m \leq m + 1$. Note that $w_1 \ldots w_m$ is a distinct parsing of $u_1 u_2 \ldots u_n$. Moreover none of these w_is are null. Thus, $m \leq m^*(u^n) - 1$ and $l_m \leq m^*(u^n)$.

(d) $1 + A + \cdots + A^{k-1}$ is the number of all possible A-ary words of length $< k$. As this is less than $m/2$, at least $m/2$ of the w_i's must have length k or more.

(e) The solution for x in $(A^x - 1)/(A - 1) = m/2$ is the quantity inside the floor. Thus $k = \lfloor x \rfloor$ satisfies the condition in (d).

(f) If $m^* < 2A\sqrt{n}$, then for sufficiently large n, it will be smaller than $4n/\log_A n$ as $\sqrt{n}\log_A n/n \to 0$. If $m^* \geq 2A\sqrt{n}$, then $n \geq \frac{1}{2}m^* \log_A \frac{m^*}{2A} \geq \frac{1}{2}m^* \log_A \frac{2A\sqrt{n}}{2A} = \frac{1}{4}m^* \log_A n$. A rearrangement gives $4n/\log_A n \geq m^*$.