ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 16 Information Theory and Coding
Take-Home Midterm Solution Oct. 27, 2020

3 problems, 4 points for each sub-problem (76 points in total).

Please submit your answer as a single PDF file on Moodle before 23h55, November 1st, 2020. Multiple submissions are allowed, only the last submission will be graded.

You can choose to scan your handwritten answer or to typeset your answer using Latex.

PROBLEM 1. A code $c : \mathcal{U} \to \{0,1\}^*$ is called *prefix-complement-free* if for every $u \in \mathcal{U}$, neither $c(u)$ nor the complement of $c(u)$ is a prefix of any other code word. (The complement of a binary string is obtained by replacing 0's with 1's and vice versa; e.g., complement of 01001 is 10110.)

a) Show that if a code $c : \mathcal{U} \to \{0,1\}^*$ is *prefix-complement-free*, then

$$\sum_{u \in \mathcal{U}} 2^{-\operatorname{length}(c(u))} \leq \frac{1}{2}.$$

Suppose $c : \mathcal{U} \to \{0,1\}^*$ is a prefix-complement-free code. Consider a code $c' : \mathcal{U} \times \{0,1\} \to \{0,1\}^*$ with twice as many codewords, whose codewords $c'((u,0))$ is equal to $c(u)$ and $c'((u,1))$ is the complement of $c(u)$. Then since c is prefix-complement-free, c' is prefix-free, so

$$1 \geq \sum_{b \in \{0,1\}} \sum_{u \in \mathcal{U}} 2^{-\operatorname{length}(c'((u,b)))} = 2 \sum_{u \in \mathcal{U}} 2^{-\operatorname{length}(c(u))},$$

which implies the required inequality.

b) Show that if $\ell : \mathcal{U} \to \{1, 2, \dots\}$ fulfills

$$\sum_{u \in \mathcal{U}} 2^{-\ell(u)} \leq \frac{1}{2}$$

then there exists a *prefix-complement-free* code $c : \mathcal{U} \to \{0,1\}^*$ with $\operatorname{length}(c(u)) = \ell(u)$ for all $u \in \mathcal{U}$.

As $\sum_{u \in \mathcal{U}} 2^{-(\ell(u)-1)} \leq 1$, we can build a prefix-free code c' such that $\operatorname{length}(c'(u)) = \ell(u) - 1$. Now define a code $c(u) = 0||c'(u)$, each codeword in this code does not contains its complement and it inherits the property of prefix-free-ness from c'.

c) Is there an operation $t : \{0,1\}^* \to \{0,1\}^*$ with $\operatorname{length}(t(s)) = \operatorname{length}(s) + 1$ such that whenever $c : \mathcal{U} \to \{0,1\}^*$ is *prefix-free*, then $\tilde{c} : \mathcal{U} \to \{0,1\}^*$ defined by $\tilde{c}(u) = t(c(u))$ is *prefix-complement-free*?

We can use the same extension method as in 1.b., namely that $t(c(u)) = 0||c(u)$.

We call a code c a *fix-complement-free* code if for every $u \in \mathcal{U}$, neither $c(u)$ nor the complement of $c(u)$ is a prefix or a suffix of any other code word. Consider these following propositions.

PROPOSITION A. If $\ell : \mathcal{U} \to \{1, 2, \dots\}$ fulfills

$$\sum_{u \in \mathcal{U}} 2^{-\ell(u)} \leq \kappa$$

then there exists a *fix-complement-free* code $c : \mathcal{U} \to \{0,1\}^*$ with $\operatorname{length}(c(u)) = \ell(u)$ for all $u \in \mathcal{U}$.

PROPOSITION B. If a code $c : \mathcal{U} \to \{0,1\}^*$ is *fix-complement-free*, then

$$\sum_{u \in \mathcal{U}} 2^{-\operatorname{length}(c(u))} \leq \kappa.$$

d) Let κ_1 be the largest value of κ such that proposition A is true. Let κ_2 be the largest value of κ such that proposition B is true. Show that $\kappa_1 < 1/2$ and $\kappa_2 \geq 1/2$.

[Hint: Give a counterexample to proposition A if $\kappa_1 \geq 1/2$, and a counterexample to proposition B if $\kappa_2 < 1/2$.]

With $\mathcal{U} = \{1, 2, 3\}$ and $\ell(1) = 2$, $\ell(2) = l(3) = 3$, we have $\sum_{u \in \mathcal{U}} 2^{-\ell(u)} = 1/2$. We claim that no choice of codewords $c(1), c(2)$ and $c(3)$ of these lengths can be fix-complement free. To see this note that there are only four choices of $c(1) : 00, 10, 01,$ and 11. We consider these four cases in turn. If $c(1) = 00$, then, by the fix-complement-free condition, $c(2)$ and $c(3)$ cannot be $000, 001, 110, 111, 100, 011$. Thus $c(2)$ and $c(3)$ have to be either 010 or 101. But these are complements of each other and thus are not valid choices for $c(2)$ and $c(3)$. The other possibilities of $c(1)$ also lead to contradiction in exactly the same way. We thus see that $\kappa_1 < 1/2$.

Furthermore, note that with $cU = 1, 2$ and $c(1) = 00, c(2) = 01$ the code c is fix-complement free and has $\sum_{u \in \mathcal{U}} 2^{-\text{ length}(c(u))} = 1/2$. Thus $\kappa_2 \geq 1/2$.

e) Find a value of κ for which you can prove proposition A, and give this proof.

We give a proof of Proposition A for $\kappa = 1/4$. Label the elements of $\mathcal{U}$ as $\mathcal{U} = u_1, \ldots, u_K$ so that $\ell(u_1) \leq \cdots \leq \ell(u_K)$. Now consider the algorithm:

(a) Mark all binary sequences $\{0, 1\}^*$ as being 'available'.

(b) For $i = 1, \ldots, K$

A1 Pick $c(u_i)$ as any available binary sequence of length $\ell(u_i)$.

A2 Mark all sequences that have $c(u_i)$ or the complement of $c(u_i)$ as either its suffix or its prefix as 'not available'.

It is clear that if the algorithm reaches step A2 it has constructed a code that is *fix-complement free*. It is also clear that the only way the algorithm may fail is in step A1, i.e., if all sequence of length $\ell(u_i)$ are marked unavailable so there is no way to pick $c(u_i)$. But the number of binary sequences of length ℓ that are available when step A1 is executed is

$$2^\ell - \sum_{j=1}^{i-1} 4.2^{\ell - \ell(u_j)} = 2^\ell \left[1 - 4 \sum_{j=1}^{i-1} 2^{-l(u_j)} \right]$$

since we start with all 2^ℓ sequences of length ℓ available and in each previous step j we eliminate $2.2^{\ell - \ell(u_j)}$ sequences that have $c(u_j)$ as its prefix or suffix; and $2^{\ell - \ell(u_j)}$ sequences that have the complement of $c(u_j)$ as its prefix or suffix.

Since $\sum_{j=1}^{i-1} 2^{-l(u_j)} < \sum_{j=1}^{K} 2^{-l(u_j)}$, the condition $\sum_{j=1}^{K} 2^{-l(u_j)} \leq 1/4$ guarantees that the number binary sequences available at step A1 is positive, so the algorithm cannot fail at step A1.

PROBLEM 2. Suppose k is a positive integer and T is a random variable that is uniformly distributed on the set $\{0/k, 1/k, \ldots, k/k\}$.

Further suppose $X_1, \ldots, X_n$ are binary random variables with

$$P_{X^n|T}(x^n \,|\, t) = \prod_{i=1}^{n} P_{X_i|T}(x_i \,|\, t),$$

and

$$P_{X_i|T}(1 \,|\, t) = t, \quad P_{X_i|T}(0 \,|\, t) = 1 - t.$$

In other words, conditional on $T = t$, the random variables $X_1, \ldots, X_n$ are i.i.d. Bernoulli(t).

Consider the following estimator $\hat{T}$ of T from the observation X^n: first compute $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then round $\bar{X}$ to the closest multiple of $1/k$.

a) Show that for $a > 0$, $\Pr\left(\left|\bar{X} - t\right| \geq a \,|\, T = t\right) \leq \frac{1}{4na^2}$. [Hint : Use Chebyshev's inequality.]

Conditional on $T = t, X_i$ are i.i.d. Bernoulli(t), in particular

$$E[X_i|T = t] = E[X_i^2|T = t] = t,$$

so conditional on $T = t, X_i$ are i.i.d. random variable with expectation t and variance $t - t^2 = t(1 - t) \leq 1/4$. Consequently, conditional on $T = t$, $\bar{X}$ has expectation t and variance $\leq 1/4n$. Using Chebyshev's inequality "for any random variable X and any $a > 0, \Pr(|X - E[X]| \geq a) \leq Var(X)/a^2)$," we immediately get the required inequality.

b) Show that $H\left(T \,|\, X^n\right) \leq 1 + p_{n,k} \log_2 k$, where $p_{n,k} = k^2/n$.

Using 2.a), and choosing $a = 1/(2k)$, we observe that $\Pr(\hat{X} \neq X) \leq p_{n,k} = k^2/n$. Now, with $p = \Pr(\hat{X} \neq X)$,

$$\begin{aligned}
H(T|X^n) &= H(T|X^n, \hat{T}) && \text{[since } \hat{T} \text{ is a function of } X^n\text{]} \\
&\leq H(T|\hat{T}) && \text{[conditioning reduced entropy]} \\
&\leq h_2(p) + p \log_2 k && \text{[Fano's inequality.]} \\
&\leq 1 + p \log_2 k && [h_2(p) \leq 1] \\
&\leq 1 + p_{n,k} \log_2 k
\end{aligned}$$

c) Show that whenever $n \geq 2k^2$, $I(T; X^n) \geq \frac{1}{2} \log_2 k - 1$.

If $n \geq 2k^2, p_{n,k} <= 1/2$, and thus

$$\begin{aligned}
I(T; X^n) &= H(T) - H(T|X^n) \\
&= \log_2(k+1) - H(T|X^n) \\
&\geq \log_2(k+1) - \frac{1}{2}\log_2(k) - 1 && \text{[Use 2.b.]} \\
&\geq \frac{1}{2}\log_2 k - 1
\end{aligned}$$

Let $\mathcal{V}$ and $\mathcal{U}$ be discrete sets and suppose for each v in $\mathcal{V}$ we are given a probability distribution on $\mathcal{U}$, denoted by $P_{U|V=v}$. Let P_V be a probability distribution on V. Together with the $P_{U|V=v}$ above, this will define the joint distribution $P_{UV}(u, v) = P_V(v)P_{U|V=v}(u)$, and thus the value of $I(U; V)$. Let Q_U be a probability distribution on $\mathcal{U}$.

d) Show that $\sum_{v\in\mathcal{V}} P_V(v) D\left(p_{U|V=v} \,\middle\|\, Q_U\right) \geq I(U;V)$.

We have,

$$\begin{aligned}
&= \sum_v P_V(v) D(P_{U|V=v}||Q_U) - I(U;V) \\
&= \sum_v P_V(v) \sum_u P_{U|V}(u|v) \left[\log_2 \frac{P_{U|V}(u|v)}{Q_U(u)} - \log_2 \frac{P_{U|V}(u|v)}{P_U(u)}\right] \\
&= \sum_{u,v} P_{UV}(uv) \log_2 \frac{P_U(u)}{Q_U(u)} \\
&= \sum_u P_U(u) \log_2 \frac{P_U(u)}{Q_U(u)} \\
&= D(P_U||Q_U) \\
&\geq 0.
\end{aligned}$$

e) Show that $\max_v D\left(P_{U|V=v} \,\middle\|\, Q_U\right) \geq C$ where $C = \max_{P_V} I(U;V)$.

We have,

$$\max_v D\left(P_{U|V=v} \,\middle\|\, Q_U\right) \geq \sum_v P_V(v) D\left(P_{U|V=v} \,\middle\|\, Q_U\right),$$

and from 2.d. we get

$$\sum_v P_V(v) D\left(P_{U|V=v} \,\middle\|\, Q_U\right) \geq I(U;V).$$

Since the left hand side does not depend on P_V and inequality holds for every P_V, we find

$$\max_v D\left(P_{U|V=v} \,\middle\|\, Q_U\right) \geq C.$$

Suppose $c : \mathcal{U} \to \{0,1\}^*$ is a uniquely decodable code.

f) Show that

$$\max_v \left\{E\left[\text{length}(c(U)) \,|\, V = v\right] - H(U|V=v)\right\} \geq C.$$

Let $Q_U(u) = \frac{2^{-\text{ length } c(u)}}{\sum_{u'} 2^{-\text{ length } c(u')}}$. Since c is uniquely decodable, $\sum_{u'} 2^{-\text{length}\, c(u')} \leq 1$; thus $\text{length}(c(u)) \geq -\log_2 Q(u)$. We now see

$$\begin{aligned}
E[\text{length}(c(U))|V=v)] - H(U|V=v) &\geq \sum_u P_{U|V}(u|v) \log \frac{1}{Q(u)} - \sum_u P(u|v) \log \frac{1}{P_{U|V}(u|v)} \\
&= D(P_{U|V=v}||Q_U).
\end{aligned}$$

Using 2.e., the required inequality follows.

Suppose we know that $X_1, \ldots, X_n$ are i.i.d. binary random variables, but we are unaware of the value $t = P(X_1 = 1)$. Despite our lack of knowledge of the true value of t, we design a uniquely decodable code

$$c_n : \{0,1\}^n \to \{0,1\}^*,$$

with the aim to make $E[\text{length}\, c_n(X^n)]$ close to $H(X^n) = nh_2(t)$.

g) No matter how we designed c_n, show that

$$\max_{t\in[0,1]}\big\{E[\text{length}\, c_n(X^n)] - nh_2(t)\big\} \geq \frac{1}{2}\log_2\lfloor\sqrt{n/2}\rfloor - 1.$$

Let $k = \lfloor\sqrt{n/2}\rfloor$ and let T be uniformly distributed in $\{0/k, \ldots, k/k\}$ as in 2.a. Note that with this choice of k we have $n \geq 2k^2$. So that $I(T; X^n) \geq \frac{1}{2}\log_2 k - 1$ from 2.c. With $U := X^n$, and $V := T$, we have

$$\begin{aligned}
\max_{t\in[0,1]}\big\{E[\text{length}\, c_n(X^n)] - nh_2(t)\big\} &\geq \max_{t\in\{0/k,\ldots,k/k\}}\big\{E[\text{length}\, c_n(X^n)] - nh_2(t)\big\} \\
&\geq C \\
&\geq I(X^n; T) \\
&\geq \frac{1}{2}\log_2 k - 1.
\end{aligned}$$

PROBLEM 3. Suppose $\ldots, U_{-1}, U_0, U_1, U_2, \ldots$ is a stationary process. Given a $k > 0$ and a function $f : \mathcal{U}^k \to \{0,1\}$, define the binary process $\ldots, V_{-1}, V_0, V_1, V_2, \ldots$ via $V_i = f(U_i, U_{i-1}, \ldots, U_{i-(k-1)})$. (E.g., $V_i = \mathbb{1}\{U_i > (U_{i-1} + \cdots + U_{i-7})/7\}$ indicates that the process exceeded its '7 day moving average' at time i, $V_i = \mathbb{1}\{U_i > U_{i-1} > ... > U_{i-(k-1)}\}$ indicates that the process has had k successive increases.)

a) Show that $Z_i = (U_i, V_i)$ and $Z'_i = (U_i, V_{i+1})$ are stationary processes.

Let $F(u, u_1, \ldots, u_k) = (u, f(u_1, \ldots, u_k))$. With this notation

$$Z_i = F(U_i, U_i, U_{i-1}, \ldots, U_{i-(k-1)}).$$

Observe now that

$$Z_1, Z_2, \cdots = F(U_1, U_1, \ldots, U_{2-k}), F(U_2, U_2, \ldots, U_{2-(k-1)}), \ldots \qquad (*)$$

and

$$Z_{n+1}, Z_{n+1}, \ldots = F(U_{n+1}, U_{n+1}, \ldots, U_{n+2-k}), F(U_{n+2}, U_{n+2}, \ldots, U_{n+3-k}), \ldots \quad (**)$$

The stationarity of U_i means that for any n, if we replace ..., $U_{-1}, U_0, U_1, \ldots$ with $\ldots, U_{n-1}, U_n, U_{n+1}$ in (*) the statistics of the left hand side will not change. But this replacement changes (*) into (**). Thus we see that for any n the statistics of $Z_1, Z_2, ...$ is the same as $Z_{n+1}, Z_{n+2}, ...$, which means that Z_i is a stationary process.

To prove the stationarity of Z'_i we can use the same reasoning above by noting that

$$Z'_1, Z'_2, ... = F(U_1, U_2, U_1, .., U_{3-k}), F(U_2, U_3, U_2, ..., U_{4-k}), ...$$

and

$$Z'_{n+1}, Z'_{n+2}, ... = F(U_{n+1}, U_{n+2}, U_{n+1}..., U_{n+3-k}), F(U_{n+2}, U_{n+3}, U_{n+2}, ..., U_{n+4-k}), ...$$

b) What is the relationship between the entropy rates of Z, Z', and U?

Observe that $H(Z^n) = H(U^n, V^n) = H(U^n) + H(V^n|U^n)$.

As V_i is a function of $U_i, ..U_{i-(k-1)}$, for $n \geq k$, we see that conditioned on U^n, the values of $V_k, ..V_n$ are fully determined, so $H(V^n|U^n) = H(V^{k-1}|U^n) \leq H(V^{k-1}) \leq (k-1)$. Thus

$$\frac{1}{n}H(U^n) \leq \frac{1}{n}H(Z_n) \leq \frac{1}{n}H(U^n) + \frac{k-1}{n}.$$

Taking the limit as n to infinity, we see that the entropy rate of U and Z are the same. The same reasoning shows that the entropy rate of U and Z' are also the same; so the three entropy rates are all equal.

c) Let $\eta_n = H(V_n \,|\, U_1, \ldots, U_{n-1})$, $n = 1, 2, \ldots$. Show that $\eta = \lim_{n\to\infty} \eta_n$ exists and $\eta \in [0, 1]$.

Note that

$$\begin{aligned}
\eta_{n+1} &= H(V_{n+1}|U_1, ..., U_n) \\
&\leq H(V_{n+1}|U_2, ..., U_n) && \text{[conditioning reduces entropy]} \\
&= H(V_n|U_1...U_{n-1}) && \text{[stationarity]} \\
&= \eta_n.
\end{aligned}$$

So $\{\eta_n\}$ is a non-increasing sequence. it is also bounded from below (by zero), consequently $\eta = \lim_{n\to\infty} \eta_n$ exists and is non-negative. Also note that $\eta_n \leq H(V_n) \leq 1$, so $\eta \leq 1$.

d) Show that the entropy rate of V is larger than or equal to η.

Let $a_n = H(V_n|V^{n-1})$. Note that the entropy rate of V is limit of the sequence a_n. For $n > k, U^{n-1}$ determines the values of $V_k, ..., V_{n-1}$. Thus

$$\begin{aligned}
\eta_n &= H(V_n|U^{n-1}) \\
&= H(V_n|U^{n-1}, V_k, ..., V_{n-1}) \\
&\geq H(V_n|V_k, ..., V_{n-1}) && \text{[conditioning reduces entropy]} \\
&= H(V_{n-k+1}|V_1, .., V_{n-k}) && \text{[stationarity]} \\
&= a_{n-(k-1)}.
\end{aligned}$$

Taking the limit of both sides over n establishes the result.

e) Explain what (if anything) is wrong with the following reasoning:
"For $n > k$, V_n is a function of $U_n, \ldots, U_{n-(k-1)}$, so

$$H(V_n \,|\, U_1, \ldots, U_{n-1}) = H(V_n \,|\, U_{n-(k-1)}, \ldots, U_{n-1}).$$

By stationarity $H(V_n \,|\, U_{n-(k-1)}, \ldots, U_{n-1}) = H(V_k \,|\, U_1, \ldots, U_{k-1})$. So $\eta_n = \eta_k$ for $n > k$; consequently $\eta = \eta_k$."

The argument is faulty because V_n being a function of $U_n ... U_{n-(k-1)}$ does not mean that it has no statistical dependence on U_i with $i < n - (k-1)$. So in the expression $H(V_n|U_1...., U_{n-1})$ one cannot remove U_i's with $i < n - (k-1)$ and claim the value of $H(V_n|U_1...., U_{n-1})$ is not changed. As an example, let U_i be a binary stationary Markov process, with $Pr(U_{i+1} = u|U_i = u) = 0.9$ for all i and all $u \in \{0, 1\}$. The stationary distribution of this process is $P(U_i = 0) = P(U_i = 1) = 1/2$. Let $k = 1$ and $V_n = U_n$. The faulty reasoning would say "V_n is only a function of U_n, so

$$H(V_n|U_1, ..., U_{n-1}) = H(V_n).$$

But the left hand side is $h_2(0.9)$ while the right hand side is 1.

Suppose we attempt to guess the value of V_n from the observations $U_1, ..., U_{n-1}$. To that end, we construct functions $g_n : \mathcal{U}^{n-1} \to \{0, 1\}$ and form $\hat{V}_n = g_n(U_1, ..., U_{n-1}),\ n = 1, 2, \ldots$. Let $p_n = \Pr(\hat{V}_n \neq V_n)$.

f) Show that $h_2(p_n) \geq \eta_n$.

We have,

$$\begin{aligned}
\eta_n &= H(V_n|U^{n-1}) \\
&= H(V_n|U^{n-1}, \hat{V}_n) && [\hat{V}_n \text{ is a function of } U^{n-1}] \\
&\leq H(V_n|\hat{V}_n) && \text{[conditioning reduces entropy]} \\
&\leq h_2(p_n) + p_n \log_2(|cV| - 1) && \text{[Fano's Inequality]} \\
&= h_2(p_n). && [|\mathcal{V}| = 2].
\end{aligned}$$

g) Let p_n^* denote the minimum possible value of p_n among all choices of g_n. Show that $p^* = \lim_{n\to\infty} p_n^*$ exists, and $h_2(p^*) \geq \eta$.

Let $g_n^*(u_1, ..., u_{n-1})$ be a function that minimizes p_n. Take now g_{n+1} as the function $g_{n+1}(u_1, ..., u_n) = g_n^*(u_2, ..., u_n)$. With this g_{n+1}, we have

$$\begin{aligned}
p_{n+1}^* &\leq Pr(V_{n+1} \neq \hat{V}_{n+1}) && [p_{n+1}^* \text{is the minimum possible}] \\
&= Pr(V_{n+1} \neq g_{n+1}(U_1...U_n)) \\
&= Pr(V_{n+1} \neq g_n^*(U_2...U_n)) \\
&= Pr(V_n \neq g_n^*(U_1...U_{n+1})) && [\text{stationarity}] \\
&= p_n^*
\end{aligned}$$

So p_n^* is a non-increasing sequence, and is bounded (by zero and one). Thus $p^* = \lim_n p^* n$ exists. Moreover, since $h_2(.)$ is continuous

$$h_2(p^*) = \lim_n h_2(p_n^*) \geq \lim_n \eta_n = \eta.$$