ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 27 Information Theory and Coding
Solution to Graded Homework Dec 1, 2020

Problem 1. Under some mobile subscription plans, you will not be charged if you hang up your call before the receiving person takes up your call. Considering this, Alice and Bob realize that Alice can send information to Bob for free by always hanging up after the first ring, and adapting the times between her calls. E.g., long wait between calls for '0', short wait for '1'.

Allowing for n calls, we can formalize this problem where Alice wants to communicate a message $M \in \mathcal{M}_n$. Assume that the messages are uniformly distributed. Let us denote the time when Alice hangs up her i-th call as T_i. We consider a family of encoding function $enc_1, enc_2, \dots, enc_n$. For the i-th call, $enc_i : \mathcal{M}_n \times \mathbb{R}_+^{i-1} \to \mathbb{R}_+$ maps the message and the previous hang up times to Alice's waiting time for the i-th call, C_i. Notice that the waiting times are required to be non-negative. Furthermore, Alice is allowed to adapt her waiting time depending on the previous hang up times.

We assume that there is a random delay between the time when Alice starts her call and the time she hears the first ring at every call. The random delays, denoted by D_i's, are i.i.d. and exponentially distributed with parameter λ, i.e., they have a probability density function given by $f(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$. For each i, the delays $D_i, \dots, D_n$ are independent of everything in the past, i.e., (M, C^i, D^{i-1}). We may also note that $E[D_i] = 1/\lambda$.

These times are related according to

$$T_i = C_i + D_i + T_{i-1}, \qquad \text{for } i \geq 1 \text{ and with } T_0 = 0.$$

Knowing $T^n = (T_1, \dots, T_n)$, Bob tries to decode the message sent with a decoder $dec : \mathbb{R}_+^n \to \mathcal{M}_n$. Furthermore, Bob does not want to wait forever, so we insist that $E[T_n] < nL$. This will ensure that in the long run the communication rate will be at least $\frac{1}{nL} \log_2 |\mathcal{M}_n|$ bits per unit time. The error probability is then expressed as

$$P_n := \Pr\big(dec(T^n) \neq M\big).$$

a) Show the following inequality

$$I(M; T^n) \leq \sum_{i=1}^{n} I(D_i + C_i; C_i).$$

By chain rule, we have

$$I(M; T^n) = \sum_{i=1}^{n} I(M; T_i | T^{i-1}) = \sum_{i=1}^{n} h(C_i + D_i | T^{i-1}) - h(C_i + D_i | M, T^{i-1})$$

As conditioning reduces entropy, we have

$$I(M; T^n) \leq \sum_{i=1}^{n} h(C_i + D_i) - h(C_i + D_i | M, T^{i-1})$$

Notice that C_i is a deterministic function of M, T^{i-1}, and D_i is independent of M,C_i and T^{i-1}, therefore

$$h(C_i + D_i|M, T^{i-1}) = h(D_i|M, T^{i-1}) = h(D_i) = h(D_i|C_i) = h(D_i + C_i|C_i).$$

which implies

$$I(M; T^n) \leq \sum_{i=1}^{n} h(C_i + D_i) - h(C_i + D_i|C_i) = \sum_{i=1}^{n} I(C_i + D_i; C_i)$$

b) Consider non-negative random variables $A_1, A_2, \ldots, A_n$, and B, where A_i is independent of B for all i. If $\frac{1}{n}\sum_{i=1}^{n} E[A_i] = \mu$, show that there exists a non-negative random variable A, independent of B, with $E[A] = \mu$ such that

$$\frac{1}{n}\sum_{i=1}^{n} I(A_i; A_i + B) \leq I(A; A + B).$$

Define an auxiliary random variable U uniformly distributed over $\{1, \ldots, n\}$. Define the random variables (A, U) such that $P_{A|U=u} = P_{A_u}$, this implies that A is non-negative and

$$E[A] = \frac{1}{n}\sum_{i=1}^{n} E[A|U = i] = \frac{1}{n}\sum_{i=1}^{n} E[A_i] = \mu.$$

Furthermore, we have that

$$\frac{1}{n}\sum_{i=1}^{n} H(A_i) = \frac{1}{n}\sum_{i=1}^{n} H(A|U = i) = H(A|U) \leq H(A)$$

and

$$\begin{aligned} I(A; A + B) &= H(A) - H(A + B|A) \\ &\geq \frac{1}{n}\sum_{i=1}^{n} H(A_i) - H(B) \\ &\geq \frac{1}{n}\sum_{i=1}^{n} I(A_i; A_i + B). \end{aligned}$$

c) Argue that there's no code sequence ensuring $P_n \to 0$ if

$$\lim_{n\to\infty} \frac{1}{n} \log_2(|\mathcal{M}_n|) > \max_{C:E[C]\leq L-\frac{1}{\lambda}} I(C; D + C), \qquad D \sim Exp(\lambda).$$

We have

$$E[T_n] = \sum_{i=1}^{n} E[C_i] + E[D_i]$$

so we have $E[T_n] \leq nL$ implies that $\sum_{i=1}^{n} E[C_i] \leq L - \frac{1}{\lambda}$.

Lets consider the problem for a specific n. Any decoder that process the received T^n into a decoded message $\hat{M}_n$ can be described by a distribution $P_{\hat{M}_n|T^n}$,such that

$M_n - T^n - \hat{M}_n$ forms a Markov chain. The data processing inequality combined with the result of a and b. implies that

$$I(\hat{M}_n; M_n) \leq I(M_n; T^n) \leq \sum_{i=1}^{n} I(D_i + C_i; C_i) \leq n \max_{C:E[C]\leq L-\frac{1}{\lambda}} I(C; D+C)$$

Therefore, we have

$$\begin{aligned} H(M_n|\hat{M}_n) &\geq H(M_n) - n \max_{C:E[C]\leq L-\frac{1}{\lambda}} I(C; D+C) \\ &= \log_2(|\mathcal{M}_n|) - n \max_{C:E[C]\leq L-\frac{1}{\lambda}} I(C; D+C). \end{aligned}$$

Finally, we know from the Fano inequality that

$$H(M_n|\hat{M}_n) \leq h_2(P_n) + P_n \log_2(|\mathcal{M}_n|).$$

The condition that we give in c. implies that there exists ϵ such that $H(M_n|\hat{M}_n) > n\epsilon$ if $n > n_0$ for a certain n_0. By Fano Inequality, this implies that $P_n \geq \frac{\epsilon - \frac{1}{n}}{\frac{1}{n}\log_2(|\mathcal{M}_n|)}$, therefore P_n cannot decay to 0.

d) What is the probability distribution of C that maximizes the mutual information in (c) ?

We have,

$$I(C; D+C) = h(D+C) - h(D)$$

From the domain of maximization, we know that $E[D+C] \leq L$. For non-negative A with $E[A] = L'$, we know that the entropy is maximized by $Exp(1/L')$. This implies that if C^* is the maximizer, then $D + C^*$ must be distributed as $Exp(1/L)$ to maximize entropy. We have for all imaginary s,

$$E[\exp(s(D+C^*))] = E[\exp(sD)]E[\exp(sC)].$$

As we know the distribution of $D + C^*$ and D, we have

$$E[\exp(sC^*)] = \frac{(1 - s/\lambda)}{(1 - sL)} = \frac{1}{\lambda L} + \left(1 - \frac{1}{\lambda L}\right)\frac{1}{(1 - sL)}.$$

As this characteristic function uniquely identify the distribution, we can determine the distribution of C^*, which is given by a mixture of a degenerate distribution at 0 with probability $1/(\lambda L)$ and an $Exp(1/L)$ with probability $1 - 1/(\lambda L)$.

e) If we want to maximize the communication rate, we can consider the following optimization problem

$$\max_{C} \frac{I(C; D+C)}{E[D+C]} = \max_{L\geq 1/\lambda} \frac{1}{L} \max_{C:E[C]=L-1/\lambda} I(C; D+C).$$

Find L that maximizes the communication rate.

From $d.$, we can compute the result of inner maximization as

$$I(C^*; D+C^*) = h(D+C^*) - h(D) = \frac{1}{\log 2}(1 + \log(L) - 1 + \log(\lambda)) = \frac{\log(L\lambda)}{\log 2}.$$

Therefore we only have to solve

$$\frac{1}{\log 2} \max_{L \geq 1/\lambda} \frac{\log \lambda L}{L}.$$

Taking the derivative, we have $f'(L) = \frac{1 - \log \lambda L}{L^2}$ from which we obtain a stationary point at $L^* = \frac{e}{\lambda}$. Furthermore, note that we have $1 - \log \lambda L < 0$ at $L > L^*$ and $1 - \log \lambda L > 0$ at $L < L^*$. Hence L^* must be a unique maximizer for $L > 1/\lambda$.

Problem 2. Suppose $\mathcal{U}$ and $\mathcal{V}$ are finite sets and we are given a conditional distribution $p_{V|U}$ (i.e., a channel).

We are given an encoder $enc : \{1, \dots, M\} \to \mathcal{U}$ with M messages. (Note that in the terminology of the course, this encoder has block length 1, and rate $\log_2 M$).

For the given encoder we consider a decoder *dec* that operates as follows. For $v \in \mathcal{V}$, for each $1 \leq m \leq M$, the decoder computes $score(m, v) = p_{V|U}(v|enc(m))$, and sets

$$dec(v) = \operatorname*{argmax}_{m \in \{1,\dots,M\}} \; score(m, v)$$

if the maximizing m is unique. If the maximizer is not unique, it chooses $dec(v)$ uniformly at random among the maximizers. [This decoder is known as the maximum likelihood decoder.]

For $m, m' \in \{1, \dots, M\}$ and $v \in \mathcal{V}$, let $Z(m, m', v) = \mathbb{1}\{score(m', v) \geq score(m, v)\}$, and let $Z(m, v) = \sum_{m' \neq m} Z(m, m', v)$ be the number of other messages that have as high score as message m.

a) Let W be a random variable taking values in $\{1, \dots, M\}$. Suppose W is the message given to the encoder for transmission. The pair (W, V) denoting the message and channel output then has distribution $p_W(w) p_{V|U}(v|enc(w))$.

   Show that for any $\rho \geq 0$, $\Pr(dec(V) \neq W) \leq E[Z(W, V)^\rho]$.

   The decoder will fail to decode the message if there are other messages with better score, in other words

$$\begin{aligned} \Pr(dec(V) \neq W) &= E[\mathbb{1}\{dec(V) \neq W\}] \\ &= E[\mathbb{1}\{Z(W, V) \geq 1\}] \\ &\leq E[Z(W, V)^\rho] \end{aligned}$$

   The last inequality is due to observation that for $x > 0$, $\mathbb{1}\{x > 1\} \leq x^\rho$ for all $\rho > 0$.

Suppose we construct the encoder randomly (denote this random encoder ENC) by choosing $ENC(1), \dots, ENC(M)$ independently, each with probability distribution p_U. With W as in (b) denoting the message to be transmitted and V denoting the channel output, we see that the distribution of $(W, ENC(1), \dots, ENC(M), V)$ given by

$$\Pr\big((W, ENC(1), \dots, ENC(M), V) = (w, u_1, \dots, u_M, v)\big) = p_W(w) p_U(u_1) \dots p_U(u_M) p_{V|U}(v|u_w).$$

b) For any $s \geq 0$ and $m \neq m'$, for the random encoder above, show that

$$E[Z(m, m', v) \,|\, W = m, ENC(m) = u, V = v] \leq p_{V|U}(v|u)^{-s} \sum_{u'} p_U(u') p_{V|U}(v|u')^s$$

   [Hint : note that, $Z(m, m', v) \leq \left[\frac{p_{V|U}(v \,|\, enc(m'))}{p_{V|U}(v \,|\, enc(m))}\right]^s$.]

   As the codeword assigned to m' is distributed according to P_U independent of the codeword for m, and the hint, we have the variable of interest as

$$\begin{aligned} &= E[Z(m, m', v) \,|\, W = m, ENC(m) = u, V = v] \\ &\leq E\left[\left(\frac{p_{V|U}(v \,|\, u')}{p_{V|U}(v \,|\, u)}\right)^s \,|\, W = m, ENC(m) = u, V = v\right] \\ &= \sum_{u'} P_U(u') \left(\frac{p_{V|U}(v \,|\, u')}{p_{V|U}(v \,|\, u)}\right)^s. \end{aligned}$$

c) For $s \geq 0$ and $0 \leq \rho \leq 1$, show that

$$E[Z(m,v)^\rho \,|\, W = m, ENC(m) = u, V = v] \leq (M-1)^\rho p_{V|U}(v|u)^{-s\rho}\Big[\sum_{u'} p_U(u')p_{V|U}(v|u')^s\Big]^\rho$$

[Hint : for $0 \leq \rho \leq 1$ and non-negative random variable A, $E[A^\rho] \leq E[A]^\rho$.]

Taking up the hint, we have the variable of interest

$$\begin{aligned}
&= E[Z(m,v)^\rho \,|\, W = m, ENC(m) = u, V = v] \\
&\leq E[Z(m,v) \,|\, W = m, ENC(m) = u, V = v]^\rho \\
&= \left(\sum_{m \neq m'} E[Z(m,m',v) \,|\, W = m, ENC(m) = u, V = v]\right)^\rho \\
&\leq \left(\sum_{m \neq m'} p_{V|U}(v|u)^{-s} \sum_{u'} p_U(u')p_{V|U}(v|u')^s\right)^\rho \\
&= (M-1)^\rho p_{V|U}(v|u)^{-s\rho}\Big[\sum_{u'} p_U(u')p_{V|U}(v|u')^s\Big]^\rho
\end{aligned}$$

d) For $s \geq 0$ and $0 \leq \rho \leq 1$, show that

$$E[Z(W,V)^\rho] \leq (M-1)^\rho \sum_v \Big[\sum_u p_U(u)p_{V|U}(v|u)^{1-s\rho}\Big]\Big[\sum_{u'} p_U(u')p_{V|U}(v|u')^s\Big]^\rho$$

We have,

$$\begin{aligned}
E[Z(W,V)^\rho] &= \sum_{u,v} P_U(u)P_{V|U}(v|u)E[Z(m,v)^\rho \,|\, W = m, ENC(m) = u, V = v] \\
&\leq (M-1)^\rho \sum_{u,v} P_U(u)P_{V|U}(v|u)^{1-s\rho}\left[\sum_{u'} P_U(u')P_{V|U}(v|u)^s\right]^\rho \\
&= (M-1)^\rho \sum_v \left[\sum_u P_U(u)P_{V|U}(v|u)^{1-s\rho}\right]\left[\sum_{u'} P_U(u')P_{V|U}(v|u)^s\right]^\rho
\end{aligned}$$

e) Show that for $0 \leq \rho \leq 1$ and $s = 1/(1+\rho)$

$$E[Z(W,V)]^\rho] \leq (M-1)^\rho \sum_v \Big[\sum_u p_U(u)p_{V|U}(v|u)^{\frac{1}{1+\rho}}\Big]^{1+\rho}$$

[Hint : with $s = 1/(1+\rho)$, note that $1 - s\rho = s$. What can you say about the sum over u and the sum over u' in (d)?]

We accept the hint and by plugging $s = 1/(1+\rho)$, we have

$$\begin{aligned}
E[Z(W,V)^\rho] &\leq (M-1)^\rho \sum_v \left[\sum_u P_U(u)P_{V|U}(v|u)^{1/(1+\rho)}\right]\left[\sum_{u'} P_U(u')P_{V|U}(v|u)^{1/(1+\rho)}\right]^\rho \\
&= (M-1)^\rho \sum_v \left[\sum_u P_U(u)P_{V|U}(v|u)^{1/(1+\rho)}\right]^{1+\rho}
\end{aligned}$$

f) Now set $U = X^n, V = Y^n$, $p_U(x^n) = \prod_i p_X(x_i)$, $p_{V|U}(y^n|x^n) = \prod_i p_{Y|X}(y_i|x_i)$, $M = \lceil 2^{nR} \rceil$. Show that for any $\rho \in [0,1]$,

$$\Pr(dec(V) \neq W) \leq E[Z(W,V)^\rho] \leq 2^{-n[G_\rho - \rho R]}$$

where

$$G_\rho = -\log_2\Big(\sum_y \Big[\sum_x p_X(x) p_{Y|X}(y|x)^{1/(1+\rho)}\Big]^{1+\rho}\Big)$$

Note that if $M = \lceil 2^{nR} \rceil$ then $M - 1 \leq 2^{nR}$. Combining our results so far gives us

$$\begin{aligned}
\Pr(dec(V) \neq W) &\leq E[Z(W,V)^\rho] \\
&\leq (M-1)^\rho \sum_{y^n} \Big[\sum_{x^n} p_{X^n}(x^n) p_{Y^n|X^n}(y^n|x^n)^{\frac{1}{1+\rho}}\Big]^{1+\rho} \\
&\leq 2^{\rho n R} \sum_{y^n} \Big[\sum_{x^n} p_{X^n}(x^n) p_{Y^n|X^n}(y^n|x^n)^{\frac{1}{1+\rho}}\Big]^{1+\rho} \\
&= 2^{\rho n R} \Bigg(\sum_{y} \Big[\sum_{x} p_{X^n}(x^n) p_{Y^n|X^n}(y^n|x^n)^{\frac{1}{1+\rho}}\Big]^{1+\rho}\Bigg)^n \\
&= 2^{-n(G_\rho - \rho R)}
\end{aligned}$$

Note that the conversion from sum over product space y^n to product over sum is only possible because both the channel is memoryless and each X_i's in the codeword is taken i.i.d.

g) Consider the Binary Erasure Channel, i.e., $BEC(q)$ with $X \in \{0,1\}$ and $Y \in \{0,?,1\}$, defined as

$$p_{Y|X}(y|x) = \begin{cases} 1-q & y = x, \\ q & y = ?, \\ 0 & \text{otherwise.} \end{cases}$$

Show that if X is uniform on $\{0,1\}$, then for this channel,

$$\lim_{\rho \to 0} \frac{G_\rho}{\rho} = 1 - q.$$

We are interested in

$$\lim_{\rho \to 0} \frac{-\log_2\Big(\sum_y \Big[\sum_x p_X(x) p_{Y|X}(y|x)^{1/(1+\rho)}\Big]^{1+\rho}\Big)}{\rho}.$$

By plugging the form of P_X and $P_{Y|X}$ we obtain,

$$\lim_{\rho \to 0} \frac{-\log_2\Big(2\Big[\frac{1}{2}(1-q)^{1/(1+\rho)}\Big]^{1+\rho} + q\Big)}{\rho} = \lim_{\rho \to 0} \frac{-\log_2\Big(2^{-\rho}(1-q) + q\Big)}{\rho}.$$

As both the numerator and denumerator is differentiable for $\rho > 0$ and both goes to 0 as ρ goes to 0, we can use the L'Hopital rule to obtain,

$$\lim_{\rho \to 0} \frac{-\log_2\Big(2^{-\rho}(1-q) + q\Big)}{\rho} = \lim_{\rho \to 0} \frac{2^{-\rho}(1-q)}{2^{-\rho}(1-q) + q} = 1 - q.$$