ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 31 — Information Theory and Coding

Homework 13 — Dec. 18, 2023

PROBLEM 1. Suppose U is $\{0,1\}$ valued with $\mathbb{P}(U=0)=\mathbb{P}(U=1)=1/2$. Suppose we have a distortion measure d given by

$$d(u,v)=\begin{cases}0, & u=v\\ 1, & (u,v)=(1,0)\\ \infty, & (u,v)=(0,1)\end{cases}$$

i.e., we never want to represent a 0 with a 1. Find $R(D)$.

PROBLEM 2. Suppose $\mathcal{U}=\mathcal{V}$ are additive groups with group operation $\oplus$. (E.g., $\mathcal{U}=\mathcal{V}=\{0,\dots,K-1\}$, with modulo K addition.) Suppose the distortion measure $d(u,v)$ depends only on the difference between u and v and is given by $g(u\ominus v)$. Let $\phi(D)$ denote $\max\{H(Z):E[g(Z)]\le D\}$.

a) Show that $\phi(D)$ is concave.

b) Let (U,V) be such that $E[d(U,V)]\le D$. Show that $I(U;V)\ge H(U)-\phi(D)$ by justifying

$$I(U;V)=H(U)-H(U|V)=H(U)-H(U\ominus V|V)\ge H(U)-H(U\ominus V)\ge H(U)-\phi(D).$$

c) Show that $R(D)\ge H(U)-\phi(D)$.

d) Assume now that U is uniform on $\mathcal{U}$. Show that $R(D)=H(U)-\phi(D)$.

PROBLEM 3. Suppose $\mathcal{U}=\mathcal{V}=\mathbb{R}$, the set of real numbers, and $d(u,v)=(u-v)^2$.

(a) Show that for any U with variance σ^2, $R(D)$ satisfies

$$h(U)-\tfrac{1}{2}\log(2\pi eD)\le R(D).$$

(b) Show that $R(D)$ does not depend on the mean of U.

Now, assume without loss of generality that U is zero-mean. Suppose we have access to a noisy observation V of U through the channel $U+Z=V$, where $Z\sim\mathcal{N}(0,\sigma_Z^2)$ and independent of U. We reconstruct U via a linear estimator $\hat{U}=aV+b$.

(c) Show that $a=\frac{\sigma^2}{\sigma^2+\sigma_Z^2}$ and $b=0$ minimizes $E[(U-\hat{U})^2]$ and for such choice of a, b, $E[(U-\hat{U})^2]=\sigma^2\frac{\sigma_Z^2}{\sigma^2+\sigma_Z^2}$.

(d) For the channel above, show that

$$I(U;V)\le\tfrac{1}{2}\log\Big(1+\tfrac{\sigma^2}{\sigma_Z^2}\Big)$$

(e) Show that for $D\le\sigma^2$

$$R(D)\le\tfrac{1}{2}\log(\sigma^2/D).$$

[Hint: Use the channel above for a candidate $p_{V|U}$.]

PROBLEM 4. Given finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, a distribution p_{XY}, $0 < \epsilon < \epsilon'$, and a sequence $x^n \in T(n, p_X, \epsilon)$, consider a random vector Y^n with independent components with $\Pr(Y_i = y) = p_{Y|X}(y|x_i)$.

For $x \in \mathcal{X}$, let $J(x) = \{i : x_i = x\}$. For an $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, let $N(x, y) = \sum_i \mathbb{1}\{(x_i, Y_i) = (x, y)\} = \sum_{i \in J(x)} \mathbb{1}\{Y_i = y\}$.

(a) Show that for each x and y, $np(x, y)(1 - \epsilon) \le E[N(x, y)] \le np(x, y)(1 + \epsilon)$, and $\mathrm{Var}(N(x, y))$ is at most n. [Hint: don't forget that x^n is in $T(n, p_X, \epsilon)$.]

(b) Show that for each x and y, both $\Pr\big(N(x, y) < np(x, y)(1 - \epsilon')\big)$ and $\Pr\big(N(x, y) > np(x, y)(1 + \epsilon')\big)$ approach to zero as n gets large. Would this be true if we had not assumed $\epsilon < \epsilon'$?

(c) Using (a) and (b) show that $\Pr\big((x^n, Y^n) \notin T(n, p_{XY}, \epsilon')\big)$ approaches 0 as gets large.

(d) Suppose now we have a distribution $p(u, x, y)$ where $p(y|ux) = p(y|x)$. [In other words, U, X, Y form a Markov chain.] Suppose (u^n, x^n) is in $T(n, p_{UX}, \epsilon)$, and Y^n has independent components as above. What can we say about $\Pr\big((u^n, x^n, Y^n) \in T(n, p_{UXY}, \epsilon')\big)$?

PROBLEM 5. Consider a two-way communication system where two parties communicate via a *common* output they both can observe and influence. Denote the common output by Y, and the signals emitted by the two parties by x_1 and x_2 respectively. Let $p(y|x_1, x_2)$ model the memoryless channel through which the two parties influence the output.

We will consider feedback-free block codes, i.e., we will use encoding and decoding functions of the form

$$\begin{aligned} &\mathrm{enc}_1\colon \{1, \dots, 2^{nR_1}\} \to \mathcal{X}_1^n && \mathrm{dec}_1\colon \mathcal{Y}^n \times \{1, \dots, 2^{nR_1}\} \to \{1, \dots, 2^{nR_2}\} \\ &\mathrm{enc}_2\colon \{1, \dots, 2^{nR_2}\} \to \mathcal{X}_2^n && \mathrm{dec}_2\colon \mathcal{Y}^n \times \{1, \dots, 2^{nR_2}\} \to \{1, \dots, 2^{nR_1}\} \end{aligned}$$

with which the parties encode their own message and decode the other party's messages. (Note that when a party is decoding the other party's message, it can make use of the knowledge of its own message).

We will say that the rate pair (R_1, R_2) is achievable, if for any $\epsilon > 0$, there exist encoders and decoders with the above form for which the average error probability is less than ϵ.

Consider the following 'random coding' method to construct the encoders:

(i) Choose probability distributions p_j on $\mathcal{X}_j$, $j = 1, 2$.

(ii) Choose $\{\mathrm{enc}_1(m_1)_i : m_1 = 1, \dots, 2^{nR_1}, i = 1, \dots, n\}$ i.i.d., each having distribution as p_1. Similarly, choose $\{\mathrm{enc}_2(m_2)_i : m_2 = 1, \dots, 2^{nR_2}, i = 1, \dots, n\}$ i.i.d., each having distribution as p_2, independently of the choices for enc_1.

For the decoders we will use typicality decoders:

(i) Set $p(x_1, x_2, y) = p_1(x_1)p_2(x_2)p(y|x_1, x_2)$. Choose a small $\epsilon > 0$ and consider the set T of ϵ-typical (x_1^n, x_2^n, y^n)'s with respect to p.

(ii) For decoder 1: given y^n and the correct m_1, dec_1 will declare $\hat{m}_2$ if it is the unique m_2 for which $(\mathrm{enc}_1(m_1), \mathrm{enc}_2(m_2), y^n) \in T$. If there is no such m_2, dec_1 outputs 0. (Similar description applies to Decoder 2.)

(a) Given that m_1 and m_2 are the transmitted messages, show that $(\mathrm{enc}_1(m_1), \mathrm{enc}_2(m_2), Y^n) \in T$ with high probability.

(b) Given that m_1 and m_2 are the transmitted messages, and $\tilde{m}_1 \neq m_1$ what is the probability distribution of $(\mathrm{enc}_1(\tilde{m}_1), \mathrm{enc}(m_2), Y^n)$?

(c) Under the assumptions in (b) show that the

$$\Pr\{(\mathrm{enc}_1(\tilde{m}_1), \mathrm{enc}_2(m_2), Y^n) \in T\} \doteq 2^{-nI(X_1;X_2Y)}.$$

(d) Show that all rate pairs satisfying

$$R_1 \leq I(X_1; YX_2), \quad R_2 \leq I(X_2; YX_1)$$

for some $p(x_1, x_2) = p(x_1)p(x_2)$ are achievable.

(e) For the case when X_1, X_2, Y are all binary and Y is the product of X_1 and X_2, show that the achievable region is strictly larger than what we can obtain by 'half duplex communication' (i.e., the set of rates that satisfy $R_1 + R_2 \leq 1$.)