

# ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

## Handout 33

Solutions to Homework 13

Information Theory and Coding

Dec. 29, 2017

PROBLEM 1. As we should never represent a 0 with a 1, we are restricted to conditional distributions with  $p_{V|U}(1|0) = 0$ . Consequently, the possible  $p_{V|U}$  are of the type

$$p_{V|U}(0|0) = 1 \quad p_{V|U}(1|0) = 0, \quad p_{V|U}(0|1) = \alpha \quad p_{V|U}(1|1) = 1 - \alpha,$$

and parametrized by  $\alpha \in [0, 1]$ . For  $p_{V|U}$  as above, we have  $\Pr(V = 1) = \frac{1}{2}(1 - \alpha)$ , and

$$E[d(U, V)] = \sum_{u,v} p_U(u) p_{V|U}(v|u) d(u, v) = \alpha/2,$$

$$I(U; V) = H(V) - H(V|U) = h_2\left(\frac{1}{2}(1 - \alpha)\right) - \frac{1}{2}h_2(\alpha) =: f(\alpha).$$

Thus  $R(D) = \min\{f(\alpha) : 0 \leq \alpha \leq \min\{1, 2D\}\}$ , with  $f(\alpha) = h_2\left(\frac{1}{2}(1 - \alpha)\right) - \frac{1}{2}h_2(\alpha)$ . It is not difficult to check that  $f$  is a decreasing function on the interval  $[0, 1]$ , and thus consequently

$$R(D) = \begin{cases} h_2\left(\frac{1}{2} - D\right) - \frac{1}{2}h_2(2D), & 0 \leq D < \frac{1}{2} \\ 0, & D \geq \frac{1}{2}. \end{cases}$$

Note that for  $D \geq \frac{1}{2}$  we can represent any  $u$  with a constant, namely  $v = 0$ , with average distortion  $1/2$ .

PROBLEM 2.

- (a) Given  $D_1, D_2$  and  $0 \leq \lambda \leq 1$  we need to show that  $\phi(D) \geq \lambda\phi(D_1) + (1 - \lambda)\phi(D_2)$ . Suppose  $p_{Z_1^*}$  and  $p_{Z_2^*}$  be the distributions on  $Z$  that achieve the maximization that define  $\phi$  for  $D_1$  and  $D_2$ , namely,  $\phi(D_1) = H(Z_1^*)$  and  $\phi(D_2) = H(Z_2^*)$  with  $E[g(Z_1^*)] \leq D_1$  and  $E[g(Z_2^*)] \leq D_2$ . Consider now the distribution  $p_{Z^*} = \lambda p_{Z_1^*} + (1 - \lambda)p_{Z_2^*}$ . For  $Z^*$  having this distribution

$$\begin{aligned} E[g(Z^*)] &= \sum_z p_{Z^*}(z)g(z) = \lambda \sum_z p_{Z_1^*}(z)g(z) + (1 - \lambda) \sum_z p_{Z_2^*}(z)g(z) \\ &= \lambda E[g(Z_1^*)] + (1 - \lambda)E[g(Z_2^*)] \leq \lambda D_1 + (1 - \lambda)D_2 = D, \end{aligned}$$

and because of the concavity of  $H$ ,  $H(Z^*) \geq \lambda H(Z_1^*) + (1 - \lambda)H(Z_2^*) = \lambda\phi(D_1) + (1 - \lambda)\phi(D_2)$ . As  $\phi(D)$  is the maximum of  $H(Z)$  over all  $Z$  with  $E[g(Z)] \leq D$ ,  $\phi(D) \geq H(Z^*)$ .

- (b) In the (in)equalities

$$\begin{aligned} I(U; V) &\stackrel{(b1)}{=} H(U) - H(U|V) \\ &\stackrel{(b2)}{=} H(U) - H(U \oplus V|V) \\ &\stackrel{(b3)}{\geq} H(U) - H(U \oplus V) \\ &\stackrel{(b4)}{\geq} H(U) - \phi(D) \end{aligned}$$

(b1) is by definition of mutual information, (b2) because for a given  $V$ ,  $U$  and  $U \oplus V$  are in one-to-one correspondence, (b3) because conditioning reduces entropy and (b4) because  $Z = U \oplus V$  has  $E[g(Z)] \leq D$ .

(c) As  $R(D) = \min\{I(U; V) : E[d(U, V)] \leq D\}$ , and by (b) for any  $U, V$  with  $E[d(U, V)] \leq D$  we have  $I(U; V) \geq H(U) - \phi(D)$ , the conclusion follows.

(d) Let  $Z$  be independent of  $U$  and have a distribution that achieves  $\phi(D)$ . Set  $V = U \oplus Z$ . Now,

$$p_{Z,V}(z, v) = p_{Z,U}(z, z \oplus v) = p_Z(z)p_U(z \oplus v) = p_Z(z)/|\mathcal{U}|.$$

By summing over  $z$  we see that  $V$  is uniformly distributed, and also that  $V$  is independent of  $Z = U \oplus V$ . Observe that the only inequalities in (b) were in (b3) and (b4), but in this case they are both equalities: (b3) because of the independence of  $Z = U \oplus V$  and  $V$ , and (b4) because  $H(Z) = \phi(D)$ .

**PROBLEM 3.** Suppose  $U, V$  satisfy  $E[(U - V)^2] \leq D$ , and set  $Z = U - V$ . As  $E[Z^2] \leq D$ , we know  $h(Z) \leq \frac{1}{2} \log(2\pi e D)$ . Also,

$$I(U; V) = h(U) - h(U|V) = h(U) - h(Z|U) \geq h(U) - h(Z) \geq h(U) - \frac{1}{2} \log(2\pi e D),$$

and consequently  $R(D) \geq h(U) - \frac{1}{2} \log(2\pi e D)$ . We now turn to the upper bound on  $R(D)$ . Assume without loss of generality that  $E[U] = 0$  so that  $\sigma^2 = E[U^2]$ . If  $D \geq \sigma^2$ , we can take  $V = 0$  for which  $E[(U - V)^2] = \sigma^2 \leq D$ , and  $I(U; V) = 0$ , so that  $R(D) = 0$ . So, we need to only consider the case  $D < \sigma^2$ . For such  $D$ , let  $Z$  be a zero mean Gaussian independent of  $U$  with variance  $D(1 - D/\sigma^2)$  and set  $V = (1 - D/\sigma^2)U + Z$ . We will show that for this choice of  $V$  we have  $E[(V - U)^2] = D$  and  $I(U; V) \leq \frac{1}{2} \log(\sigma^2/D)$ , which will then establish that  $R(D) \leq \frac{1}{2} \log(\sigma^2/D)$ . To that end observe that  $V - U = -(D/\sigma^2)U + Z$  and thus  $E[(V - U)^2] = (D/\sigma^2)^2 E[U^2] + E[Z^2] = D$ . Turning our attention now to  $I(U; V)$ , first compute  $E[V^2] = (1 - D/\sigma^2)^2 E[U^2] + E[Z^2] = \sigma^2 - D$ , so  $h(V) \leq \frac{1}{2} \log(2\pi e(\sigma^2 - D))$ . Furthermore,

$$\begin{aligned} h(V|U) &= h(V - (1 - D/\sigma^2)U | U) = h(Z|U) = h(Z) \\ &= \frac{1}{2} \log(2\pi e \text{Var}(Z)) = \frac{1}{2} \log(2\pi e D(1 - D/\sigma^2)). \end{aligned}$$

Thus

$$I(U; V) = h(V) - h(V|U) \leq \frac{1}{2} \log \frac{\sigma^2 - D}{D(1 - D/\sigma^2)} = \frac{1}{2} \log(\sigma^2/D).$$

**PROBLEM 4.**

- (a) Since the channel is memoryless and feedback-free transmission is assumed, from code construction, it is obvious that  $(\text{enc}_1(m_1), \text{enc}_2(m_2), Y^n)$  is an i.i.d. length- $n$  sequence of  $(X_1, X_2, Y)$ 's drawn from distribution  $p(x_1, x_2, y) = p_1(x_1)p_2(x_2)p(y|x_1, x_2)$ . Therefore, for sufficiently large  $n$ , the probability of this sequence being  $\epsilon$ -typical is as high as desired.
- (b) Now,  $(\text{enc}_1(\tilde{m}_1), \text{enc}_2(m_2), Y^n)$  is an i.i.d. sequence (of length  $n$ ) whose components are distributed according to  $p_1(x_1)p(y, x_2)$  where  $p(y, x_2) = \sum_{x'_1} p_1(x'_1)p_2(x_2)p(y|x'_1, x_2)$ .
- (c)  $\Pr\{(\text{enc}_1(\tilde{m}_1), \text{enc}_2(m_2), Y^n) \in T\}$  is the probability of a length  $n$  i.i.d. sequence  $X_1^n$  whose elements have distribution  $p_1$  being jointly  $\epsilon$ -typical (with respect to the distribution  $p_1(x_1)p(y, x_2|x_1)$  where  $p(y, x_2|x_1) = p(x_2)p(y|x_1, x_2)$ ) with an independent length  $n$  sequence of  $(X_2, Y)^n$  whose elements have distribution  $p(y, x_2)$  (defined in (b)). Thus, as we have seen in the course,

$$\Pr\{(\text{enc}_1(\tilde{m}_1), \text{enc}_2(m_2), Y^n) \in T\} \doteq 2^{-nI(X_1, X_2, Y)}.$$

(In the course we have seen this result for two random variables  $X$  and  $Y$ ; it is obvious that we can replace  $X$  by  $X_1$  and  $Y$  by  $(X_2, Y)$  to derive the desired result).

- (d) From (a) we know that the probability of the correct message  $m_1$  not being on the list of typical  $m_1$ 's at decoder 2 is small, say at most  $\epsilon/2$ .

From (c), the probability of each incorrect  $\tilde{m}_1$  being on that list (at decoder 2) is equal (up to sub-exponential factors) to  $2^{-nI(X_1;X_2Y)}$ . Since there are  $M - 1 \leq 2^{nR_1}$  such  $\tilde{m}_1$ 's, the probability of having *an* incorrect message on the list is, by the union bound, at most  $2^{n[R_1 - I(X_1;X_2Y)]}$  which is exponentially small in  $n$  provided that  $R_1 < I(X_1;X_2Y)$ . Thus, for large enough  $n$ , this probability is also smaller than  $\epsilon/2$ .

Consequently, the average probability of decoding error at decoder 2 is at most  $\epsilon$  provided that  $R_1 < I(X_1, X_2Y)$ .

By symmetry, the average probability of decoding error at decoder 1 is smaller than  $\epsilon$  if  $R_2 < I(X_2, X_1, Y)$ .

Since the average probability of error (over the generation of codebooks) is small (for rate pairs  $(R_1, R_2)$  satisfying  $R_1 < I(X_1; Y, X_2)$  and  $R_2 < I(X_2; Y, X_1)$ ), there exist a pair of codebooks of rates  $(R_1, R_2)$  in the ensemble for which the average error probability is small, thus such  $(R_1, R_2)$ 's are achievable.

- (e) Firstly note that since  $X_1$  and  $X_2$  are independent,  $I(X_1; YX_2) = I(X_1; Y|X_2)$  (similarly  $I(X_2; YX_1) = I(X_2; Y|X_1)$ ).

Since  $Y = X_1 \times X_2$ , conditioned on  $\{X_2 = 0\}$ ,  $Y$  contains no information about  $X_1$ , whereas conditioned on  $\{X_2 = 1\}$ ,  $Y = X_1$ . Assuming  $\Pr\{X_1 = 1\} = p_1$  and  $\Pr\{X_2 = 1\} = p_2$ ,

$$\begin{aligned} I(X_1; Y|X_2) &= \Pr\{X_2 = 0\}I(X_1; Y|X_2 = 0) + \Pr\{X_2 = 1\}I(X_1; Y|X_2 = 1) \\ &= 0 + p_2h_2(p_1) \end{aligned}$$

where  $h_2(\cdot)$  is the binary entropy function. Similarly it follows that  $I(X_2; Y|X_1) = p_1h_2(p_2)$ .

Suppose  $p_1 = p_2 = p$ , then all rates  $(R_1, R_2)$  satisfying

$$R_1 < ph_2(p) \quad R_2 < ph_2(p)$$

are achievable. In particular,  $ph_2(p) \geq \frac{1}{2}$  for some  $p \geq \frac{1}{2}$  (it evaluates to  $\frac{1}{2}$  at  $p = \frac{1}{2}$  but it is increasing, so it will go above  $\frac{1}{2}$  as  $p$  increases). The set of achievable rate pairs corresponding to such  $p$ 's violate  $R_1 + R_2 < 1$ .

#### PROBLEM 5.

- (a) We know that an i.i.d. sequence of length  $m$  with each component having distribution  $q$  will belong to the typical set  $T(p, m, \delta)$  with probability  $2^{-m[D(p||q)+O(\delta)]}$ . In the setting of the problem  $m = n(u)$ ,  $q = p_V$ ,  $p = p_{V|U=u}$ . Thus, the probability we are asked for is given by  $2^{-n(u)[D(p_{V|U=u}||p_V)+O(\delta)]}$ .
- (b) Since the events  $\{V(u) \in T(p_{V|U=u}, n(u), \delta)\}_{u \in \mathcal{U}}$  are independent the probability that all of the  $|\mathcal{U}|$  events occurs is the product of the occurrence of each of them we computed in (a). Thus the probability of the joint occurrence is  $2^{-n[F+O(\delta)]}$  with

$$F = \sum_u \frac{n(u)}{n} D(p_{V|U=u}||p_V).$$

(c) When  $u^n \in T(n, p_U, \delta)$  we have  $p(u)(1 - \delta) \leq n(u)/n \leq p(u)(1 + \delta)$ . Thus,

$$F = \sum_u p(u) D(p_{V|U=u} \| p_V) + O(\delta).$$

But  $D(p_{V|U=u} \| p_V) = \sum_v p(v|u) \log \frac{p(v|u)}{p(v)}$  and thus the sum that defines  $F$  equals  $I(U; V)$ , and we conclude that the probability in (b) equals  $2^{-n[I(U; V) + O(\delta)]}$ .

(d) When  $u^n \in T(n, p_U, \delta)$  and  $V(u) \in T(n(u), p_{V|U=u}, \delta)$  for each  $u$ , then the number of occurrences of the pair  $(u, v)$  in the sequence  $(u^n, V^n)$  is upper and lower bounded by

$$n(1 + \delta)p_U(u)(1 + \delta)p_{V|U}(v|u) = n(1 + 2\delta + \delta^2)p_{UV}(u, v)$$

and

$$n(1 - \delta)p_U(u)(1 - \delta)p_{V|U}(v|u) = n(1 - (2\delta - \delta^2))p_{UV}(u, v).$$

Thus we have  $(u^n, V^n)$  belonging to  $T(n, P_{UV}, 2\delta + \delta^2)$ .

(e) With  $1 \geq \delta' \geq 3\delta$  we have  $\delta' \geq 2\delta + \delta^2$ , and thus  $T = T(n, p_{UV}, 2\delta + \delta^2) \subset T' = T(n, p_{UV}, \delta')$ . Consequently, for any  $u^n$ ,  $\Pr((u^n, V^n) \in T') \geq \Pr((u^n, V^n) \in T)$ . By (d) we know that the event  $\{(u^n, V^n) \in T\}$  includes the event “ $V(u) \in T(n(u), p_{V|U=u}, \delta)$  for every  $u$ ”. By (c), this last event has  $2^{-n[I(U; V) + O(\delta)]}$ .