Handout 24 Information Theory and Coding
Solutions to Homework 10 Nov. 26, 2024

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

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

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

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

$$R(D) = \begin{cases} h_2(\frac{1}{2} - D) - \frac{1}{2} h_2(2D), & 0 \le D < \frac{1}{2} \\ 0, & D \ge \frac{1}{2}. \end{cases}$$

Note that for $D \ge \frac{1}{2}$ 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 $0 \le \lambda \le 1$ we need to show that $\phi(D) \ge \lambda \phi(D_1) + (1 - \lambda) \phi(D_2)$. Suppose $p_{Z_1^*}$ and $p_{Z_2^*}$ be the distributions on Z that achieve the maximization that define ϕ for D_1 and D_2, namely, $\phi(D_1) = H(Z_1^*)$ and $\phi(D_2) = H(Z_2^*)$ with $E[g(Z_1^*)] \le D_1$ and $E[g(Z_2^*)] \le D_2$. Consider now the distribution $p_{Z^*} = \lambda p_{Z_1^*} + (1 - \lambda) p_{Z_2^*}$. For Z^* having this distribution

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

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

(b) In the (in)equalities

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

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

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

(d) Let Z be independent of U and have a distribution that achieves $\phi(D)$. Set $V = U \ominus Z$. 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 $Z = U \ominus V$. 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 $Z = U \ominus V$ and V, and (b4) because $H(Z) = \phi(D)$.

Problem 3. (a) $I(U;V) = h(U) - h(U|V) = h(U) - h(U-V|V) \ge h(U) - h(U-V) = h(U) - h(Z)$ where $Z := U - V$. Now let us minimize the lower bound. Since U has a fixed distribution, the problem is equivalent to maximizing $h(Z)$ under the constraint $E[Z^2] \le D$. From the lectures, we know that such distribution is a zero-mean Gaussian with variance D. Therefore we obtain

$$R(D) = \min_{\substack{P_{V|U}:\\ E[(U-V)^2]\le D}} I(U;V) \ge \min_{\substack{P_{V|U}:\\ E[(U-V)^2]\le D}} h(U) - h(Z) \ge h(U) - \tfrac{1}{2}\log(2\pi e D).$$

(b) Suppose $p^*_{V|U}$ is the distribution achieving the minimum in $R(D)$, and $E[(U-V)^2] = D^*$ for such choice of $p^*_{V|U}$. First, we prove that $E[U] = E[V]$. Suppose $E[U] = \mu_U \ne E[V] = \mu_V$ and let $\tilde{U} := U - \mu_U$, $\tilde{V} := V - \mu_V$ be centered versions of U, V. Then, $E[(U-V)^2] = E[(\tilde{U} - \mu_U - \tilde{V} + \mu_V)^2] = E[(\tilde{U} - \tilde{V})^2] + (\mu_U - \mu_V)^2$. Since shifts do not change the mutual information and thus $I(U;V) = I(U; V - \mu_V + \mu_U)$, one can always make $\mu_U = \mu_V$ and achieve a smaller distortion than $E[(U-V)^2]$ by eliminating the $(\mu_U - \mu_V)^2$ term. Hence, μ_U must be equal to μ_V. In this case, $E[(\tilde{U} - \tilde{V})^2] = D^*$ and both $I(U;V)$ and D^* does not depend on the mean of U.

(c) First, we show $b = 0$. $E[(U-\hat{U})^2] = E[(U-aV-b)^2] = E[(U-aV)^2] - 2E[(U-aV)]b + b^2 = E[(U-aV)^2] + b^2$ as both U and V are zero-mean. Hence b must be 0 to minimize $E[(U-\hat{U})^2]$. For $b = 0$, $E[(U-\hat{U})^2] = E[(U-aV)^2] = E[U^2] - 2aE[UV] + a^2E[V^2]$. Since this is a quadratic function of a, it is minimized at $a = \frac{E[UV]}{E[V^2]} = \frac{\sigma^2}{\sigma^2 + \sigma_Z^2}$ and the minimum value turns out to be $\frac{\sigma^2\sigma_Z^2}{\sigma^2+\sigma_Z^2}$.

(d) Observe that the above channel is an additive Gaussian noise channel. We know that the mutual information between the input and the output is upper bounded by $\frac{1}{2}\log\left(1 + \frac{\mathrm{Var}(U)}{\sigma_Z^2}\right)$.

(e) Let $\tilde{V} = \hat{U} = aV$, where $\hat{U}$ and a are as in part (b) and (c). Now, observe that the $E[(U-\tilde{V})^2] = \frac{\sigma^2\sigma_Z^2}{\sigma^2+\sigma_Z^2}$ and $I(U;\tilde{V}) = I(U;V) \le \frac{1}{2}\log\left(1 + \frac{\sigma^2}{\sigma_Z^2}\right) = \frac{1}{2}\log\left(\frac{\sigma^2}{E[(U-\tilde{V})^2]}\right)$. Given $D \le \sigma^2$, we can choose σ_Z^2 to ensure $E[(U-\tilde{V})^2] = D$, so $R(D) \le I(U;\tilde{V}) \le \frac{1}{2}\log\left(\frac{\sigma^2}{D}\right)$.

Problem 4. (a) Observe that for any $i \in J(x)$, $E[\mathbb{1}\{Y_i = y\}] = p_{Y|X}(y|x)$. Therefore, $E[N(x,y)] = |J(x)|p_{Y|X}(y|x)$. Since $x^n \in T(n, p_x, \epsilon)$, we have $(1-\epsilon)np(x) \le |J(x)| \le (1+\epsilon)np(x)$. Hence, $(1-\epsilon)np(x,y) \le E[N(x,y)] \le (1+\epsilon)np(x,y)$. We also have $\mathrm{Var}(N(x,y)) = \mathrm{Var}\left(\sum_{i\in J(x)} \mathbb{1}\{Y_i = y\}\right) = \sum_{i\in J(x)} \mathrm{Var}(\mathbb{1}\{Y_i = y\})$ because Y_i's are chosen i.i.d. and $\sum_{i\in J(x)} \mathrm{Var}(\mathbb{1}\{Y_i = y\}) \le |J(x)| \le n$ because we know that for a random variable that takes binary values, its variance can be at most 1.

(b) Write

$$\Pr(N(x,y) < np(x,y)(1-\epsilon')) \leq \Pr(N(x,y) - E[N(x,y)] < np(x,y)(\epsilon - \epsilon')).$$

As $\epsilon' > \epsilon$, we can apply Chebyshev's inequality to the rightmost term to obtain

$$\Pr(N(x,y) < np(x,y)(1-\epsilon')) \leq \frac{\mathrm{Var}(N(x,y))}{n^2 p(x,y)^2(\epsilon'-\epsilon)^2} \leq \frac{1}{np(x,y)^2(\epsilon'-\epsilon)^2},$$

which tends to 0 as $n \to \infty$. Proceed similarly to obtain the same result for $\Pr(N(x,y) > np(x,y)(1+\epsilon'))$.

For $\epsilon' < \epsilon$; it is not guaranteed that the above expressions tend to zero for all $x^n \in T(n, p_x, \epsilon)$. In fact, had we taken a $x^n \in T(n, p_x, \epsilon)$, but $x^n \notin T(n, p_x, \epsilon')$; we would have at least one $x \in \mathcal{X}$ such that $J(x) > (1+\epsilon')np(x)$ and $E[N(x,y)] > (1+\epsilon')np(x,y)$, which makes it impossible for $\Pr(N(x,y) > (1+\epsilon')np(x,y))$ to go to zero.

(c)

$$\begin{aligned}
&\Pr((x^n, Y^n) \notin T(n, p_{XY}, \epsilon')) \\
&\quad = \Pr(\exists x, y \in \mathcal{X} \times \mathcal{Y} : N(x,y) \notin [(1-\epsilon')np(x,y), (1+\epsilon')np(x,y)]) \\
&\quad \leq \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \left[\Pr(N(x,y) < np(x,y)(1-\epsilon')) + \Pr(N(x,y) > np(x,y)(1+\epsilon'))\right] \\
&\quad \leq \frac{1}{n} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \frac{2}{p(x,y)^2(\epsilon'-\epsilon)^2},
\end{aligned}$$

which tends to 0 as $n \to \infty$.

(d) With (U, X) playing the role of X in (a,b,c), we see that the event we ask is exactly the complement of the event in (c). Therefore, its probability goes to 1.