École Polytechnique Fédérale de Lausanne
School of Computer and Communication Sciences

Handout 24 Information Theory and Coding
Solutions to Homework 9 Dec. 25, 2019

Problem 1.

(a) This is by the definition of mutual information once we note that $p_{Y|X}(y|x) = p_Z(y-x)$.

(b) Note that $p_X(x)p_Z(y-x)$ is simply the joint distribution of (x, y), and thus the integral

$$\iint p_X(x)p_Z(y-x)\ln\frac{\mathcal{N}_{\sigma^2}(y-x)}{\mathcal{N}_{\sigma^2+P}(y)}\,dxdy.$$

is an expectation, namely

$$E\ln\frac{\mathcal{N}_{\sigma^2}(Y-X)}{\mathcal{N}_{\sigma^2+P}(Y)}.$$

Substituting the formula for $\mathcal{N}$, this in turn, is

$$\begin{aligned}
&E\ln\frac{\mathcal{N}_{\sigma^2}(Y-X)}{\mathcal{N}_{\sigma^2+P}(Y)}\\
&\quad=\frac{1}{2}\ln\left(1+P/\sigma^2\right)+\frac{1}{2(\sigma^2+P)}E[Y^2]-\frac{1}{2\sigma^2}E[(Y-X)^2]\\
&\quad=\frac{1}{2}\ln\left(1+P/\sigma^2\right)+\frac{1}{2(\sigma^2+P)}E[(X+Z)^2]-\frac{1}{2\sigma^2}E[Z^2]\\
&\quad=\frac{1}{2}\ln\left(1+P/\sigma^2\right)+\frac{1}{2(\sigma^2+P)}E[X^2+Z^2+2XZ]-\frac{1}{2}\\
&\quad=\frac{1}{2}\ln\left(1+P/\sigma^2\right)+\frac{1}{2(\sigma^2+P)}(P+\sigma^2+0)-\frac{1}{2}\\
&\quad=\frac{1}{2}\ln\left(1+P/\sigma^2\right)
\end{aligned}$$

(c) The steps we need to justify read

$$\begin{aligned}
\ln(1+P/\sigma^2)-I(X;Y) &= \iint p_X(x)p_Z(y-x)\ln\frac{\mathcal{N}_{\sigma^2}(y-x)p_Y(y)}{\mathcal{N}_{\sigma^2+P}(y)p_Z(y-x)}\,dxdy\\
&\le \iint \frac{p_X(x)\mathcal{N}_{\sigma^2}(y-x)p_Y(y)}{\mathcal{N}_{\sigma^2+P}(y)}\,dxdy-1\\
&= \int p_Y(y)\,dy-1\\
&= 0.
\end{aligned}$$

The first equality is by substitution of parts (a) and (b). The inequality is by $\ln(x) \le x-1$. The next equality is by noting that

$$\int p_X(x)\mathcal{N}_{\sigma^2}(y-x)dx = (p_X * \mathcal{N}_{\sigma^2})(y) = (\mathcal{N}_P * \mathcal{N}_{\sigma^2})(y) = \mathcal{N}_{P+\sigma^2}(y).$$

The last equality is because any density function integrates to 1.

(d) The conclusion is made by noting that the right hand side of the first equality in (c) is equal to zero if $p_Z = \mathcal{N}_{\sigma^2}$.

PROBLEM 2. Let the input distribution be p. We thus have

$$p(-1) + p(0) + p(1) = 1 \qquad p(-1) \geq 0, p(0) \geq 0, p(1) \geq 0$$

(since p is a distribution) and, to satisfy $E[b(X)] \leq \beta$ we must have

$$p(-1) + p(1) = 1 - p(0) \leq \beta.$$

Moreover,

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &\overset{(a)}{=} H(Y) - p(0) \\ &\overset{(b)}{\leq} 1 - p(0) \\ &\overset{(c)}{\leq} \min\{1, \beta\}. \end{aligned}$$

where (a) follows because given $\{X = -1\}$ or $\{X = 1\}$ there is no uncertainity in Y while given $\{X = 0\}$, Y is uniformly distributed in $\{-1, 1\}$, (b) holds since Y is binary with equality if $p(-1) + \frac{1}{2}p(0) = p(1) + \frac{1}{2}p(0) = \frac{1}{2}$ (which happens if we choose $p(1) = p(-1) = \frac{1}{2}(1 - p(0)))$ and (c) holds because of the cost constraint and is equality if we choose $p(0) = \max\{1 - \beta, 0\}$. Hence, the capacity is

$$C = \begin{cases} \beta, & \text{if } \beta \leq 1 \\ 1, & \text{if } \beta > 1 \end{cases}.$$

PROBLEM 3.

(a) All rates less than $\frac{1}{2}\log_2(1 + \frac{P}{\sigma^2})$ are achievable.

(b) The new noise $Z_1 - \rho Z_2$ has zero mean and variance $\mathrm{E}((Z_1 - \rho Z_2)^2) = \sigma^2(1 - \rho^2)$. Therefore, all rates less than $\frac{1}{2}\log_2(1 + \frac{P}{\sigma^2(1-\rho^2)})$ are achievable.

(c) The capacity is $C = \max I(X; Y_1, Y_2) = \max(h(Y_1, Y_2) - h(Z_1, Z_2)) = \frac{1}{2}\log_2(1 + \frac{P}{\sigma^2(1-\rho^2)})$. This shows that the scheme used in (b) is a way to achieve capacity.

PROBLEM 4.

(a) Since C is non-empty, it contains some codeword x. By linearity C must contain $x + x$. But, for any x, $x + x$ is the all-zero sequence since we are doing modulo-2 sums. So, C contains the all-zero sequence.

(b) The elements of D' are those sequences of the form $x + y$ where y is in D. Since x is in C and D is a subset of C, any x and y are both in C, and so is their sum.

(c) Suppose there was an element z common to D and D'. Then $z = x + y$ where y is in D. Since we assumed that D is a linear subset, then $z + y$ is also in D. But $z + y$ equals x, and we arrive at the contradiction that x is in D.

(d) Since the mapping $y \mapsto x + y$ is a bijection, D and D' are in one-to-one correspondence, and hence have the same number of elements.

(e) Suppose z_1 and z_2 are in $D \cup D'$. There are four possibilities: (1) both z_1 and z_2 are in D, (2) both z_1 and z_2 are in D', (3) z_1 is in D, z_2 is in D', (4) z_1 is in D', z_2 is in D. In case (1), the linearity of D implies that $z_1 + z_2$ is in D. In case (2), $z_1 = x + y_1$ and $z_2 = x + y_2$ for some y_1 and y_2 both in D, then $z_1 + z_2 = x + x + y_1 + y_2 = y_1 + y_2$ is in D. In case (3) $z_2 = x + y_2$ and $z_1 + z_2 = x + (z_1 + y_2)$, which is in D', and similarly in case (4). Thus in all cases $z_1 + z_2$ is in $D \cup D'$ and we see that $D \cup D'$ is a linear subset of C.

(f) We thus see that if at the beginning of step (ii) D is a linear subset of C, at the end of step (iii) $D \cup D'$ is linear, is a subset of C because both D and D' are, and has twice as many elements of D since D' has the same number of elements of D and is disjoint from it. Thus, when the algorithm terminates, D contains all elements of C and since it is a subset of C it must equal C. Furthermore, its size, being equal to successive doublings of 1, is a power of 2.

Problem 5. Let x and x' be two different codewords in the extended Hamming code. Let z and z' be the parts of x and x' that come from the Hamming code (i.e., z is all but the last bit of x, and z' that of x'), and p and p' be the bits appended to z and z' to get x and x'. Since x and x' are different then so are z and z': if $z = z'$ then $p = p'$ and x and x' would have been the same. Thus, $d_H(z, z') \geq 3$ since z and z' are different Hamming codewords. On the other hand, if $d_H(z, z') = 3$, then z and z' must have different parity: if they both had an even number of 1's or both had an odd number of 1's they would have differed in an even number of places and $d_H(z, z')$ would have been an even number. Thus, if $d_H(z, z') = 3$ then $p \neq p'$ and we have $d_H(x, x') = 4$. If $d_H(z, z') \geq 4$ then clearly $d_H(x, x') \geq 4$. We thus see that the minimum distance of the new code is 4.

Consider the following procedure to decode

> Given a sequence y, compare it to all the codewords and find the number of positions in which y differs from them. If there is a unique codeword for which this number is smallest, declare that codeword. If not, declare 'errors were detected'.

If the minimum distance d of a code is an even number, $d = 2j$, then if a sequence y differs from the transmitted codeword x by up to $j - 1$ places, then y will be close to the transmitted codeword than to any other and the decoder will correctly decode x. If however, y differs from x in j places, then no other codeword will be closer to y, but there might be a codeword x' which also differs from y in j places. In such a case the decoder will not be able to correct but detect the errors. In particular, if $d = 4$, then all single errors are corrected and all double errors are detected (may even be corrected).

Problem 6.

(a) Note first that the sum of two even-weight codewords is of even weight, the sum of two odd-weight codewords is of even weight and the sum of an odd-weight codeword with an even-weight codeword is of odd weight.

If the code contains no odd-weight codeword then we are done. Otherwise let x be an odd-weight codeword. Then the mapping $y \mapsto x + y$ is a bijection between even-weight and odd-weight codewords, and we conclude that there must be an equal number of odd-weight and even-weight codewords.

(b) The same proof above applies: either all codewords have a zero at the nth digit, or there is a codeword x with has a 1 in its nth digit. The mapping $y \mapsto x + y$ gives a bijection between codewords who have a zero at the nth digit and codewords which have a 1 at the nth digit. In the first case, when all codewords have a zero at the nth digit, one can improve the code by simply deleting the nth digit from each codeword: no matter what the message is, the same symbol would have been transmitted, giving no additional information.

(c) To find the average number of 1's per codewords, one would find the total number of 1's in all codewords, and divide this sum by the number of codewords. Suppose there are M codewords. Arrange the codewords in rows, and count the total number of 1's by going over columns one by one. Since each column contains at most $M/2$ ones, and there are N columns, the total number of 1's is less than or equal to $MN/2$. Dividing by M we see that the average number of 1's per codeword is at most $N/2$.