EE326 Introduction to Information Theory and Coding

- TA in charge of Homework #5: Minyeol Bae (willy3690@kaist.ac.kr)

7.8

Let $p_X(1) = a$. Then,

$$\begin{aligned} H(Y|X) &= \sum_{x=0}^{1} p_X(x) H(Y|X=x) \\ &= p_X(0)H(1) + p_X(1)H(1/2) = a \end{aligned}$$

From Q,

$$\begin{aligned} p_Y(0) &= p_{Y|X}(0|0)p_X(0) + p_{Y|X}(0|1)p_X(1) = 1 - a + \frac{1}{2}a = 1 - \frac{1}{2}a \\ p_Y(1) &= p_{Y|X}(1|0)p_X(0) + p_{Y|X}(1|1)p_X(1) = \frac{1}{2}a \end{aligned}$$

Then,

$$\begin{aligned} H(Y) &= H\left(\frac{1}{2}a\right) \\ I(X;Y) = H(Y) - H(Y|X) &= H\left(\frac{1}{2}a\right) - a \\ &= -\frac{1}{2}a\log\left(\frac{1}{2}a\right) - \left(1 - \frac{1}{2}a\right)\log\left(1 - \frac{1}{2}a\right) - a \end{aligned}$$

Let $f(a) = H\left(\frac{1}{2}a\right) - a = -\frac{1}{2}a\log\left(\frac{1}{2}a\right) - \left(1-\frac{1}{2}a\right)\log\left(1-\frac{1}{2}a\right) - a$.

$$\begin{aligned} f'(a) &= -\frac{1}{2}\log\left(\frac{1}{2}a\right) - \frac{1}{2} + \frac{1}{2}\log\left(1 - \frac{1}{2}a\right) + \frac{1}{2} - 1 \\ &= \frac{1}{2}\log\left(\frac{2}{a} - 1\right) - 1 \end{aligned}$$

Thus, $f'(a) = 0$ at $a = 2/5$.

$$\therefore C = \max_{p_X(x)} I(X;Y) = f(2/5) = H\left(\frac{1}{5}\right) - \frac{2}{5}, \quad \text{(bits)}$$

achieved by the input distribution

$$p_X(x) = \begin{cases} 2/5, & x = 0 \\ 3/5, & x = 1 \end{cases}$$

7.12

Without loss of generality, let $\mathcal{X} = \{x_1, x_2, x_3\}$ and $\mathcal{Y} \in \{y_1, y_2, y_3\}$.
Let $p_X(x_1) = a$, $p_X(x_2) = b$, and $p_X(x_3) = c$.
Then, $a + b + c = 1$ and $H(a, b, c) \le H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \log 3$ always holds.

$$\begin{aligned} H(Y|X) &= \sum_{x \in \{x_1, x_2, x_3\}} p_X(x) H(Y|X = x)) \\ &= aH\left(\frac{2}{3}, \frac{1}{3}, 0\right) + bH\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) + cH\left(0, \frac{1}{3}, \frac{2}{3}\right) \\ &= (1 - b)H\left(\frac{2}{3}, \frac{1}{3}, 0\right) + bH\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \\ &\ge H\left(\frac{2}{3}, \frac{1}{3}, 0\right) = \log 3 - \frac{2}{3} \end{aligned}$$

where equality holds if and only if $b = 0$.
From $P_{y|x}$,

$$p_Y(y_1) = p_{Y|X}(y_1|x_1)p_X(x_1) + p_{Y|X}(y_1|x_2)p_X(x_2) + p_{Y|X}(y_1|x_3)p_X(x_3) = \frac{2}{3}a + \frac{1}{3}b$$
$$p_Y(y_2) = p_{Y|X}(y_2|x_1)p_X(x_1) + p_{Y|X}(y_2|x_2)p_X(x_2) + p_{Y|X}(y_2|x_3)p_X(x_3) = \frac{1}{3}a + \frac{1}{3}b + \frac{1}{3}c = \frac{1}{3}$$
$$p_Y(y_3) = p_{Y|X}(y_3|x_1)p_X(x_1) + p_{Y|X}(y_3|x_2)p_X(x_2) + p_{Y|X}(y_3|x_3)p_X(x_3) = \frac{1}{3}b + \frac{2}{3}c$$

Thus,

$$\begin{aligned} H(Y) &= H\left(\frac{2}{3}a + \frac{1}{3}b, \frac{1}{3}, \frac{1}{3}b + \frac{2}{3}c\right) \\ &= H\left(\frac{1}{3}(1 + a - c), \frac{1}{3}, \frac{1}{3}(1 - (a - c))\right) \\ &\le H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \log 3 \end{aligned}$$

where equality holds if and only if $a = c$.
Thus, if $a = c = \frac{1}{2}$ and $b = 0$, $I(X;Y) = H(Y) - H(Y|X)$ is maximized.

$$\begin{aligned} C &= \max_{p_X(x)} I(X;Y) = \max_{p_X(x)} (H(Y) - H(Y|X)) \\ &= \log 3 - H\left(\frac{2}{3}, \frac{1}{3}, 0\right) = \frac{2}{3}, \quad \text{(bits)} \end{aligned}$$

achieved by the input distribution $p_X(x_1) = p_X(x_3) = \frac{1}{2}$ and $p_X(x_2) = 0$, which indicates that the second input symbol (x_2) has a zero probability.
The intuitive reasons are (a) we can achieve $H(Y) = \log 3$ without using x_2 and (b) $H(Y|X = x_2) = \log 3$, where $\log 3$ is the maximum possible value of entropy. Reason (b) means that the use of x_2 increases $H(Y|X)$. Therefore, even if there were other ways to achieve $H(Y) = \log 3$ while using x_2, such a construction would increase $H(Y|X)$, resulting in a decreased $I(X;Y)$.

7.24

$$H(Y|X) = H(X+Z|X) = H(Z) = \log 3$$

Thus, the capacity of this channel is solely dependent on $H(Y)$.

$$C = \max_{p_X(x)} I(X;Y) = \max_{p_X(x)} (H(Y) - H(Y|X)) = \max_{p_X(x)} H(Y) - \log 3$$

(a) We can simply derive following inequality.

$$H(Y) \overset{(a)}{\leq} \log|\mathcal{Y}| \overset{(b)}{\leq} \log(|\mathcal{X}||\mathcal{Z}|) = \log 12$$

where equality of (a) holds if and only if Y is uniformly distributed and equality of (b) holds if and only if all the $x+z$'s are different for all $x \in \mathcal{X}, z \in \mathcal{Z}$ (e.g., $z_1 = 0$, $z_2 = 4$, and $z_3 = 8$).
Assume that equality of (b) is achieved. Then, if X is uniformly distributed, since Z is uniformly distributed and there is no overlap in $x+z$, $\forall x \in \mathcal{X}, z \in \mathcal{Z}$, Y is also uniformly distributed, then equality of (a) is achieved. Thus,

$$C = \max_{p_X(x)} H(Y) - \log 3 = \log 12 - \log 3 = 2 \quad \text{(bits)}$$

achieved by the uniform input distribution.

(b) The minimum capacity occurs when $|\mathcal{Y}|$ is the smallest. Since $Y = X + Z$,

$$|\mathcal{Y}| = |\mathcal{X} + \mathcal{Z}| \geq |\mathcal{X}| + |\mathcal{Z}| - 1 = 4 + 3 - 1 = 6$$

where the minimum is achieved when $\mathcal{Z}$ consists of three consecutive integers. Thus, to consider the minimum capacity, without loss of generality, we select the simple case: $\mathcal{Z} = \{0, 1, 2\}$. Then, $\mathcal{Y} = \{0, 1, 2, 3, 4, 5\}$ and

$$H(Y) \leq \log|\mathcal{Y}| = \log 6$$

where equality holds if and only if Y has a uniform distribution.
Let

$$p_X(x) = \begin{cases} \frac{1}{2}, & x \in \{0, 3\} \\ 0, & \text{otherwise} \end{cases}$$

(i.e., we only use $x = 0$ and $x = 3$)
Then,

$$\begin{aligned} P(Y = y) &= P(X + Z = y | X = 0) p_X(0) + P(X + Z = y | X = 3) p_X(3) \\ &= P(Z = y) \times \frac{1}{2} + P(3 + Z = y) \times \frac{1}{2} \\ &\overset{(a)}{=} P(Z = y \vee 3 + Z = y) \times \frac{1}{2} = \frac{1}{6} \end{aligned}$$

where (a) is because $0 \leq Z < 3$. This means that Y can have a uniform distribution. Thus,

$$C = \max_{p_X(x)} H(Y) - \log 3 = \log 6 - \log 3 = 1 \quad \text{(bits)}$$

7.17

(a) The best code maximizes the minimum distance (d). A distance of $d = 3$ is only possible for a code with two codewords (e.g., 000 and 111). Since we require 4 codewords, d must be less than 3.

We can easily construct a code with $d = 2$. Any successful construction of a code with $d = 2$ in this case have a symmetrical arrangement where the distance between every pair of distinct codewords is exactly 2. Thus, without loss of generality, we consider the following code:

$$C = \{000, 101, 110, 011\}$$

The decoding rule is to select the nearest codeword. Then, the probability of success is determined as follows:

(i) correct codeword
With probability of $(1-\epsilon)^3$, the correct codeword is received. In this case, decoding is always successful.

(ii) 1-bit error
With probability of $3\epsilon(1-\epsilon)^2$, 1-bit error occurs. Since $d_{min} = 2$, the received vector is not a codeword (e.g., if 000 is sent, 001 is received). Due to the code's symmetry, this received vector always has exactly three nearest codewords (e.g., the received vector 001 is distance 1 from 000, 101, and 011). Assuming the decoder selects one of these three at random, the success rate in this specific event is 1/3.

(iii) 2 or 3-bit error
Errors of 2 or 3 bits always cause the received vector to be closer to an incorrect codeword, resulting in decoding failure.

Thus, the total probability of success is

$$P_{success} = (1-\epsilon)^3 + \frac{1}{3}3(1-\epsilon)^2 = (1-\epsilon)^2 = 0.81$$

and the probability of error for this code is

$$P_{error} = 1 - P_{success} = 1 - 0.81 = 0.19$$

(b) This configuration offers no redundancy and thus no error correction capability. The decoding rule is to select the nearest codeword. Since every received vector is a valid codeword, the decoded message is always the received vector.
Therefore, the probability of success ($P_{success}$) is the probability that zero bit errors occur

$$P_{success} = (1-\epsilon)^3 = 0.729$$

and the probability of error is

$$1 - P_{success} = 1 - 0.729 = 0.271$$

(c) By the symmetry, the decoding failure probability is the same regardless of whether 000 or 111 was transmitted. Without loss of generality, assume that 000 is transmitted. Then, the error only occurs when EEE is received and the decoder randomly guesses the incorrect codeword, 111. This probability is

$$\begin{aligned} P_{error} &= \frac{1}{2}P(\text{EEE}|000) \\ &= \frac{1}{2}\epsilon^3 = 0.0005 \end{aligned}$$

(d)

(a) Without loss of generality,

$$C = \{000, 101, 110, 011\}$$

By the symmetry, the decoding failure probability is the same regardless of the transmitted codeword. Without loss of generality, assume that 000 is transmitted. Then, consider the probability of success:

(i) 000 is received
With probability of $(1-\epsilon)^3$, 000 is received. In this case, decoding is always successful.

(ii) 1 erasure is received
With probability of $3\epsilon(1-\epsilon)^2$, 1 erasure occurs. Then, the received vector consists of two 0s. The only codeword that contains two or more 0s is 000. Therefore, the decoder can uniquely determine that 000 was the transmitted codeword.

(iii) 2 erasures are received
With probability of $3\epsilon^2(1-\epsilon)$, 2 erasures occur. For every resulting pattern (e.g., $0EE$), the received vector is consistent with two valid codewords. (e.g., 0EE fits 000 and 011). The decoder must guess, resulting in a success rate of 1/2.

(iv) 3 erasures are received
With probability of ϵ^3, 3 erasures occur. Since the received vector provides no information, the decoder randomly guesses among all four codewords. Thus, the success rate is 1/4.

Thus, the total probability of success is

$$P_{success} = (1-\epsilon)^3 + 3(1-\epsilon)^2 + \frac{1}{2}3\epsilon^2(1-\epsilon) + \frac{1}{4}\epsilon^3 = 0.98575$$

and the probability of error for this code is

$$P_{error} = 1 - P_{success} = 1 - 0.98575 = 0.01425$$

(b) Consider n erasures are received. The probability of exactly n erasures occurring is given by the binomial distribution:

$$P(n \text{ erasures}) = \binom{3}{n} \epsilon^n (1-\epsilon)^{3-n}$$

Since every sequence is a valid codeword, a received vector with n erasures is consistent with 2^n possible codewords. Since only one of these is the correct transmitted sequence, the decoding failure rate of this case is

$$P(\text{error}|n \text{ erasures}) = \frac{2^n - 1}{2^n}$$

Therefore, the probability of error for this code is

$$\begin{aligned} P_{error} &= \sum_{n=0}^{3} P(\text{error}|n \text{ erasures}) P(n \text{ erasures}) \\ &= \sum_{n=0}^{3} \frac{2^n - 1}{2^n} \binom{3}{n} \epsilon^n (1-\epsilon)^{3-n} = 0.142625 \end{aligned}$$

8.3

(a)

$$I(X;Y) = h(Y) - h(Y|X)$$

$$h(Y|X) = h(X+Z|X) = h(Z) = \log a$$

To calculate $h(Y)$, we need to calculate $f_Y(y)$.

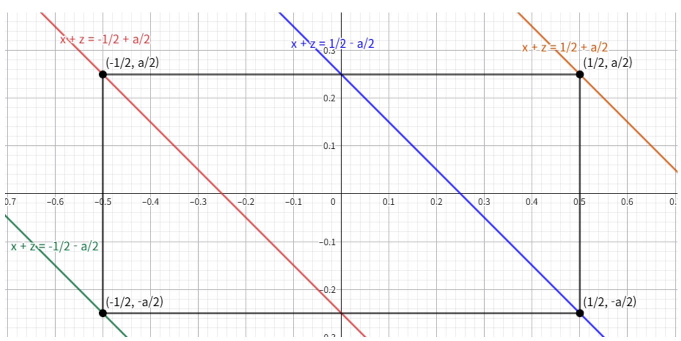
$$f_Y(y) = \frac{d}{dy}F_Y(y)$$
$$F_Y(y) = P(Y \le y) = P(X + Z \le y)$$

$$f_{X,Z}(x,z) = f_X(x)f_Z(z) = \begin{cases} \frac{1}{a}, & -\frac{1}{2} \le x \le \frac{1}{2}, -\frac{a}{2} \le z \le \frac{a}{2} \\ 0, & \text{otherwise} \end{cases}$$

Since $f_{X,Z}(x,z)$ is uniform,

$$P(X+Y \le y) = \frac{1}{a} \times \left(\text{Area of } \left\{(x,z) : x + z \le y, -\frac{1}{2} \le x \le \frac{1}{2}, -\frac{a}{2} \le z \le \frac{a}{2}\right\}\right)$$

(i) $0 \le a \le 1$



(i-i) $y \le -1/2 - a/2$

$$F_Y(y) = P(X + Z \le y) = 0$$
$$f_Y(y) = \frac{d}{dy}F_Y(y) = 0$$

(i-ii) $-1/2 - a/2 < y \leq -1/2 + a/2$

$$F_Y(y) = P(X + Z \leq y) = \frac{1}{a} \times \frac{1}{2}\left(y + \frac{1+a}{2}\right)^2$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{a}y + \frac{1+a}{2a}$$

$$-\int_{-\frac{1+a}{2}}^{\frac{-1+a}{2}} f_Y(y)\log f_Y(y)dy = -\int_{-\frac{1+a}{2}}^{\frac{-1+a}{2}} \left(\frac{1}{a}y + \frac{1+a}{2a}\right)\log\left(\frac{1}{a}y + \frac{1+a}{2a}\right)dy$$

$$= -\left[\frac{a}{2}\left(\frac{1}{a}y + \frac{1+a}{2a}\right)^2 \log\left(\frac{1}{a}y + \frac{1+a}{2a}\right)\right]_{-\frac{1+a}{2}}^{\frac{-1+a}{2}} + \int_{-\frac{1+a}{2}}^{\frac{-1+a}{2}} \frac{1}{2}\left(\frac{1}{a}y + \frac{1+a}{2a}\right)dy$$

$$= \left[\frac{a}{4}\left(\frac{1}{a}y + \frac{1+a}{2a}\right)^2\right]_{-\frac{1+a}{2}}^{\frac{-1+a}{2}} = \frac{a}{4}$$

(i-iii) $-1/2 + a/2 < y \leq 1/2 - a/2$

$$F_Y(y) = P(X + Z \leq y) = \frac{1}{a} \times \left(\frac{1}{2}a^2 + a\left(y - \frac{a-1}{2}\right)\right)$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = 1$$

$$-\int_{\frac{-1+a}{2}}^{\frac{1-a}{2}} f_Y(y)\log f_Y(y)dy = -\int_{\frac{-1+a}{2}}^{\frac{1-a}{2}} 1\log 1dy = 0$$

(i-iv) $1/2 - a/2 < y \leq 1/2 + a/2$

$$F_Y(y) = P(X + Z \leq y) = \frac{1}{a} \times \left(a - \frac{1}{2}\left(\frac{1+a}{2} - y\right)^2\right)$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = -\frac{1}{a}y + \frac{1+a}{2a}$$

$$-\int_{\frac{1-a}{2}}^{\frac{1+a}{2}} f_Y(y)\log f_Y(y)dy = -\int_{\frac{1-a}{2}}^{\frac{1+a}{2}} \left(-\frac{1}{a}y + \frac{1+a}{2a}\right)\log\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)dy$$

$$= -\left[-\frac{a}{2}\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)^2 \log\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)\right]_{\frac{1-a}{2}}^{\frac{1+a}{2}} + \int_{\frac{1-a}{2}}^{\frac{1+a}{2}} \frac{1}{2}\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)dy$$

$$= \left[-\frac{a}{4}\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)^2\right]_{\frac{1-a}{2}}^{\frac{1+a}{2}} = \frac{a}{4}$$

(i-v) $y > 1/2 + a/2$

$$F_Y(y) = P(X + Z \leq y) = a$$

$$f_Y(y) = 0$$

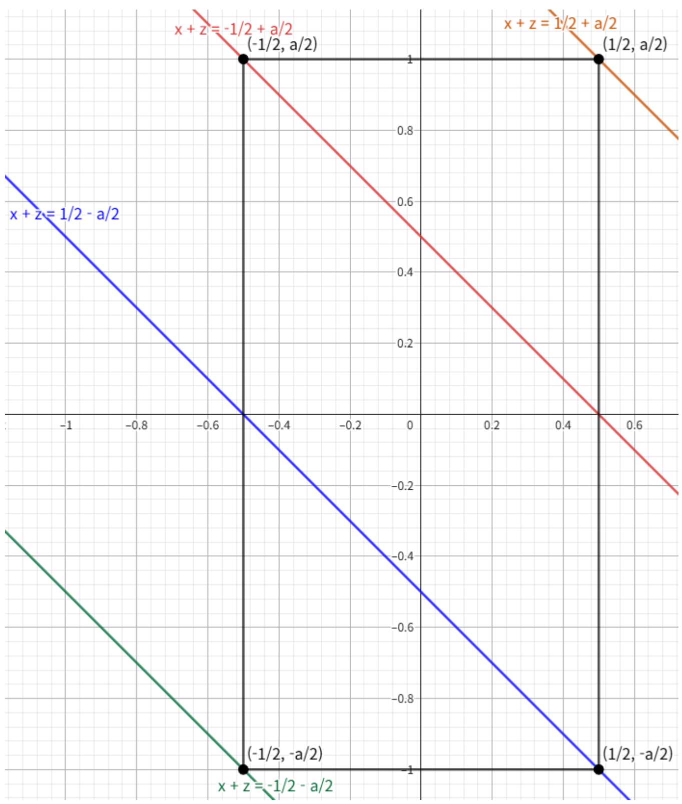
Then,

$$
\begin{aligned}
h(Y) &= -\int_{-\frac{1+a}{2}}^{\frac{1+a}{2}} f_Y(y)\log f(y)dy \\
&= -\int_{-\frac{1+a}{2}}^{\frac{-1+a}{2}} f_Y(y)\log f(y)dy - \int_{\frac{-1+a}{2}}^{\frac{1-a}{2}} f_Y(y)\log f(y)dy - \int_{\frac{1-a}{2}}^{\frac{1+a}{2}} f_Y(y)\log f(y)dy \\
&= \frac{a}{4} + 0 + \frac{a}{4} = \frac{a}{2}
\end{aligned}
$$

Thus,

$$I(X;Y) = h(Y) - h(Y|X) = \frac{a}{2} - \log a \quad \text{(nats)}$$

(ii) $a > 1$



(i-i) $y \le -1/2 - a/2$

$$F_Y(y) = P(X+Z \le y) = 0$$
$$f_Y(y) = \frac{d}{dy}F_Y(y) = 0$$

(i-ii) $-1/2 - a/2 < y \le 1/2 - a/2$

$$F_Y(y) = P(X+Z \le y) = \frac{1}{a} \times \frac{1}{2}\left(y + \frac{1+a}{2}\right)^2$$
$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{a}y + \frac{1+a}{2a}$$
$$-\int_{-\frac{1+a}{2}}^{\frac{1-a}{2}} f_Y(y)\log f_Y(y)dy = -\int_{-\frac{1+a}{2}}^{\frac{1-a}{2}} \left(\frac{1}{a}y + \frac{1+a}{2a}\right)\log\left(\frac{1}{a}y + \frac{1+a}{2a}\right)dy$$
$$= -\left[\frac{a}{2}\left(\frac{1}{a}y + \frac{1+a}{2a}\right)^2 \log\left(\frac{1}{a}y + \frac{1+a}{2a}\right)\right]_{-\frac{1+a}{2}}^{\frac{1-a}{2}} + \int_{-\frac{1+a}{2}}^{\frac{1-a}{2}} \frac{1}{2}\left(\frac{1}{a}y + \frac{1+a}{2a}\right)dy$$
$$= \frac{1}{2a}\log a + \left[\frac{a}{4}\left(\frac{1}{a}y + \frac{1+a}{2a}\right)^2\right]_{-\frac{1+a}{2}}^{\frac{1-a}{2}} = \frac{1}{2a}\log a + \frac{1}{4a}$$

(i-iii) $1/2 - a/2 < y \le -1/2 + a/2$

$$F_Y(y) = P(X+Z \le y) = \frac{1}{a} \times \left(\frac{1}{2} + 1 \times \left(y - \frac{1-a}{2}\right)\right)$$
$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{a}$$
$$-\int_{\frac{1-a}{2}}^{\frac{-1+a}{2}} f_Y(y)\log f_Y(y)dy = -\int_{\frac{1-a}{2}}^{\frac{-1+a}{2}} \frac{1}{a}\log\frac{1}{a}dy = \frac{a-1}{a}\log a$$

(i-iv) $-1/2 + a/2 < y \le 1/2 + a/2$

$$F_Y(y) = P(X+Z \le y) = \frac{1}{a} \times \left(a - \frac{1}{2}\left(\frac{1+a}{2} - y\right)^2\right)$$
$$f_Y(y) = \frac{d}{dy}F_Y(y) = -\frac{1}{a}y + \frac{1+a}{2a}$$
$$-\int_{\frac{-1+a}{2}}^{\frac{1+a}{2}} f_Y(y)\log f_Y(y)dy = -\int_{\frac{-1+a}{2}}^{\frac{1+a}{2}} \left(-\frac{1}{a}y + \frac{1+a}{2a}\right)\log\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)dy$$
$$= -\left[-\frac{a}{2}\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)^2 \log\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)\right]_{\frac{-1+a}{2}}^{\frac{1+a}{2}} + \int_{\frac{-1+a}{2}}^{\frac{1+a}{2}} \frac{1}{2}\left(-\frac{1}{a}y + \frac{1-a}{2a}\right)dy$$
$$= \frac{1}{2a}\log a + \left[-\frac{a}{4}\left(-\frac{1}{a}y + \frac{1+a}{2a}\right)^2\right]_{\frac{-1+a}{2}}^{\frac{1+a}{2}} = \frac{1}{2a}\log a + \frac{1}{4a}$$

(i-v) $y > 1/2 + a/2$

$$\begin{aligned}F_Y(y) &= P(X+Z \le y) = a \\ f_Y(y) &= 0\end{aligned}$$

Then,

$$\begin{aligned}
h(Y) &= -\int_{-\frac{1+a}{2}}^{\frac{1+a}{2}} f_Y(y)\log f(y)dy \\
&= -\int_{-\frac{1+a}{2}}^{\frac{1-a}{2}} f_Y(y)\log f(y)dy - \int_{\frac{1-a}{2}}^{\frac{-1+a}{2}} f_Y(y)\log f(y)dy - \int_{\frac{-1+a}{2}}^{\frac{1+a}{2}} f_Y(y)\log f(y)dy \\
&= \frac{1}{2a}\log a + \frac{1}{4a} + \frac{a-1}{a}\log a + \frac{1}{2a}\log a + \frac{1}{4a} = \log a + \frac{1}{2a}
\end{aligned}$$

Thus,

$$I(X;Y) = h(Y) - h(Y|X) = \log a + \frac{1}{2a} - \log a = \frac{1}{2a} \quad \text{(nats)}$$

$$\therefore I(X;Y) = \begin{cases} \frac{a}{2} - \log a, & 0 \le a \le 1 \\ \frac{1}{2a}, & a > 1 \end{cases}$$

(b)

$$C = \max_{f_X(x)} I(X;Y) = \max_{f_X(x)} (h(Y) - h(Y|X))$$

Since $a = 1$, $h(Y|X) = \log a = 0$. Then, the capacity optimization simplifies to maximizing $h(Y)$. Since $Y \in [-1, 1]$,

$$h(Y) \le \log(1 - (-1)) = \log 2$$

where equality holds if and only if Y is uniformly distributed ($\because$peak-limited). To confirm that the capacity $C = \log 2$ is achievable, we must find an input distribution $f_X(x)$ that makes Y uniformly distributed on $[-1, 1]$. Let

$$f_X(x) = \frac{1}{2}\delta\left(x + \frac{1}{2}\right) + \frac{1}{2}\delta\left(x - \frac{1}{2}\right)$$

Then, since $-\frac{1}{2} + Z \in [-1, 0]$ and $\frac{1}{2} + Z \in [0, 1]$,

$$f_Y(y) = f_Z\left(y + \frac{1}{2}\right) f_X\left(-\frac{1}{2}\right) + f_Z\left(y - \frac{1}{2}\right) f_X\left(\frac{1}{2}\right) = \frac{1}{2},\ y \in [-1, 1]$$

Thus, Y can be uniformly distributed.

$$\therefore C = \max_{f_X(x)} (h(Y) - h(Y|X)) = \log 2 - 0 = \log 2 \quad \text{(nats)}$$

achieved by the input distribution

$$f_X(x) = \frac{1}{2}\delta\left(x + \frac{1}{2}\right) + \frac{1}{2}\delta\left(x - \frac{1}{2}\right)$$

8.9

Since translation does not change the differential entropy (e.g., $h(X + c) = h(X)$), without loss of generality, $E[X] = E[Y] = E[Z] = 0$.
Let σ_x^2, σ_y^2, and σ_z^2 be the variances of X, Y, and Z, respectively. Then, the correlation coefficient ρ_{xz} between X and Z is

$$\begin{aligned}
\rho_{xz} &= \frac{E[XZ] - E[X]E[Z]}{\sigma_x \sigma_z} \\
&= \frac{E[E_Y[XZ|Y]]}{\sigma_x \sigma_z} \\
&\overset{(a)}{=} \frac{E[E_Y[X|Y][E_Y[Z|Y]]}{\sigma_x \sigma_z} \\
&\overset{(b)}{=} \frac{E[(\frac{\sigma_x \rho_1}{\sigma_Y} Y)(\frac{\sigma_z \rho_2}{\sigma_Y} Y)]}{\sigma_x \sigma_z} = \rho_1 \rho_2
\end{aligned}$$

where (a) is because of the Markov chain property and (b) is because (X, Y), (Y, Z) are jointly Gaussian.
Then, since X and Z are correlated Gaussian random variables with the correlation coefficient ρ_{xz},

$$I(X; Z) = -\frac{1}{2} \log(1 - \rho_{xz}^2) = -\frac{1}{2} \log(1 - \rho_1^2 \rho_2^2)$$