Fundamentals of Information Theory

Homework 5

Name: Zhou Shouchen
Student ID: 2021533042

Due 23:59 (CST), Dec. 3, 2024

Problem 1

8.1 Differential entropy. Evaluate the differential entropy $h(X) = -\int f \ln f$ for the following:

(a) The exponential density, $f(x) = \lambda e^{-\lambda x}, x \geq 0$.

(b) The Laplace density, $f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$.

(c) The sum of X_1 and X_2, where X_1 and X_2 are independent normal random variables with means μ_i and variances $\sigma_i^2, i = 1, 2$.

Solution

(a) $X \sim f(x) = \lambda e^{-\lambda x}$, so we have

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^{+\infty} x\left(\lambda e^{-\lambda x}\right) dx \\
&= \lambda \int_0^{+\infty} x\left(-\frac{1}{\lambda}\right) d\left(e^{-\lambda x}\right) \\
&= -xe^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\
&= 0 + \left(-\frac{1}{\lambda}\right) e^{-\lambda x}\Big|_0^{+\infty} \\
&= \frac{1}{\lambda}
\end{aligned}$$

So we can get that

$$\begin{aligned}
h(X) &= -\int_0^{+\infty} f(x)\ln f(x)dx = -\int_0^{+\infty} \lambda e^{-\lambda x} \ln\left(\lambda e^{-\lambda x}\right) dx \\
&= -\int_0^{+\infty} \lambda e^{-\lambda x}(\ln\lambda - \lambda x)dx \\
&= (-\ln\lambda)\int_0^{+\infty} \lambda e^{-\lambda x} dx + \lambda \int_0^{+\infty} x\left(\lambda e^{-\lambda x}\right) dx \\
&= -\ln\lambda + \lambda\mathbb{E}(X) \qquad \left(\int_0^{+\infty} \lambda e^{-\lambda x} dx \text{ is the integration of PDF of } X \text{ on its support, so it is 1}\right) \\
&= \ln\left(\frac{e}{\lambda}\right) \text{ nats} \\
&= \log\left(\frac{e}{\lambda}\right) \text{ bits}
\end{aligned}$$

(b) $X \sim f(x) = \dfrac{1}{2}\lambda e^{-\lambda|x|}$, so we have

$$\begin{aligned}
h(X) &= -\int_{-\infty}^{+\infty} f(x)\ln f(x)dx = -\int_{-\infty}^{+\infty} \frac{1}{2}\lambda e^{-\lambda|x|} \ln\left(\frac{1}{2}\lambda e^{-\lambda|x|}\right) dx \\
&= -2\int_0^{+\infty} \frac{1}{2}\lambda e^{-\lambda x}\left(\ln\left(\frac{\lambda}{2}\right) - \lambda x\right) dx \qquad \text{(The integrated function is a even function)} \\
&= -\ln\left(\frac{\lambda}{2}\right)\int_0^{+\infty} \underbrace{\lambda e^{-\lambda x}}_{\text{Expo(X)'s PDF}} dx + \lambda \underbrace{\int_0^{+\infty} x\left(\lambda e^{-\lambda x}\right) dx}_{\text{Expo(X)'s expectation}} \\
&= \ln\left(\frac{2}{\lambda}\right)\cdot 1 + \lambda\cdot\frac{1}{\lambda} = \ln\left(\frac{2e}{\lambda}\right) \text{ nats} \\
&= \log\left(\frac{2e}{\lambda}\right) \text{ bits}
\end{aligned}$$

(c) $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, so we have $X = X_1+X_2 \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$, i.e. $X \sim \mathcal{N}(\mu, \sigma^2)$, where $\mu = \mu_1 + \mu_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$.
So we have

$$X \sim f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

So we have the variance of X:

$$\sigma^2 = \mathbb{E}_{X\sim f(x)} = \mathbb{E}_{X\sim f(x)} \left(X - \mathbb{E}(X)\right)^2 = \mathbb{E}_{X\sim f(x)} \left(X-\mu\right)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

So we can get that

$$\begin{aligned}
h(X) &= -\int_{-\infty}^{+\infty} f(x) \ln f(x) dx \\
&= -\int_{-\infty}^{+\infty} f(x) \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) dx \\
&= \ln\left(\sqrt{2\pi}\sigma\right) \int_{-\infty}^{+\infty} f(x)dx + \frac{1}{2\sigma^2} \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \\
&= \ln\left(\sqrt{2\pi}\sigma\right) + \frac{1}{2\sigma^2}\sigma^2 \\
&= \frac{1}{2}\ln\left(2\pi\sigma^2\right) + \frac{1}{2} \\
&= \frac{1}{2}\ln\left(2\pi e\sigma^2\right) \text{ nats} \\
&= \frac{1}{2}\log\left(2\pi e\sigma^2\right) \text{ bits}
\end{aligned}$$

i.e. $h(X) = \frac{1}{2}\log\left(2\pi e\left(\sigma_1^2 + \sigma_2^2\right)\right)$ bits.

Problem 2

8.3 Uniformly distributed noise. Let the input random variable X to a channel be uniformly distributed over the interval $-\frac{1}{2} \leq x \leq +\frac{1}{2}$. Let the output of the channel be $Y = X + Z$, where the noise random variable is uniformly distributed over the interval $-a/2 \leq z \leq +a/2$.

(a) Find $I(X;Y)$ as a function of a.

(b) For $a = 1$ find the capacity of the channel when the input X is peak-limited; that is, the range of X is limited to $-\frac{1}{2} \leq x \leq +\frac{1}{2}$. What probability distribution on X maximizes the mutual information $I(X;Y)$?

(c) (Optional) Find the capacity of the channel for all values of a, again assuming that the range of X is limited to $-\frac{1}{2} \leq x \leq +\frac{1}{2}$.

Solution

From the discription, we can get that

$$X \sim \text{Unif}\left(-\frac{1}{2}, \frac{1}{2}\right), f_X(x) = 1, x \in \left[-\frac{1}{2}, \frac{1}{2}\right]$$

$$Z \sim \text{Unif}\left(-\frac{a}{2}, \frac{a}{2}\right), f_Z(z) = \frac{1}{a}, z \in \left[-\frac{a}{2}, \frac{a}{2}\right]$$

So we have:

$$f_Y(y) = \int_{-\infty}^{+\infty} \Pr\left(X = z - y | Z = z\right) \Pr(Z = z) dz = \int_{-\infty}^{+\infty} f_X(y - z) f_Z(z) dz = f_X * f_Z$$

Where $*$ denotes the convolution operation.

And since f_X, f_Z could be regarded as two rectangle signals, so their convolution if the sum of their products in the intersection range. It could be easier to compute with slicing window. And since the range of Z is $[-\frac{a}{2}, \frac{a}{2}]$, which means that $a > 0$, so we have:

1. If $a = 1$:

$$f_Y(y) = \begin{cases} y + 1, & -1 \leq y < 0 \\ 1 - y, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

2. If $0 < a < 1$:

$$f_Y(y) = \begin{cases} \frac{1}{a}\left(y + \frac{1}{2} + \frac{a}{2}\right), & -\frac{1}{2} - \frac{a}{2} \leq y < -\frac{1}{2} + \frac{a}{2} \\ 1, & -\frac{1}{2} + \frac{a}{2} \leq y < \frac{1}{2} - \frac{a}{2} \\ -\frac{1}{a}\left(y - \frac{1}{2} - \frac{a}{2}\right), & \frac{1}{2} - \frac{a}{2} \leq y \leq \frac{1}{2} + \frac{a}{2} \\ 0, & \text{otherwise} \end{cases}$$

3. If $a > 1$:

$$f_Y(y) = \begin{cases} \frac{1}{a}\left(y + \frac{1}{2} + \frac{a}{2}\right), & -\frac{1}{2} - \frac{a}{2} \leq y < \frac{1}{2} - \frac{a}{2} \\ \frac{1}{a}, & \frac{1}{2} - \frac{a}{2} \leq y < -\frac{1}{2} + \frac{a}{2} \\ -\frac{1}{a}\left(y - \frac{1}{2} - \frac{a}{2}\right), & -\frac{1}{2} + \frac{a}{2} \leq y \leq \frac{1}{2} + \frac{a}{2} \\ 0, & \text{otherwise} \end{cases}$$

And before we compute the mutual information, we calculate a mostly used calculation first:

$$\begin{aligned}
I &= \int_0^1 t\ln t \ \mathrm{dt} \\
&= [t \cdot (t\ln t)]|_0^1 - \int_0^1 t \ \mathrm{d}\,(t\ln t) \\
&= 0 - 0 - \int_0^1 t\cdot(1+\ln t) \ \mathrm{dt} \\
&= -\int_0^1 t \ \mathrm{dt} - \int_0^1 t\ln t \ \mathrm{dt} \\
&= -\frac{1}{2} - I \\
\Rightarrow I &= -\frac{1}{4}
\end{aligned}$$

i.e. we have proved that

$$I = \int_0^1 t\ln t \ \mathrm{dt} = -\frac{1}{4}$$

(a) Combined with previous discussion, we can get the mutual information:

$$\begin{aligned}
I(X;Y) &= h(Y) - h(Y|X) \\
&= h(Y) - h(Z|X) \qquad (\text{Since } Y = X+Z) \\
&= h(Y) - h(Z) \qquad (\text{Since } Z \perp X)
\end{aligned}$$

And since $Z \sim \mathrm{Unif}\left(-\frac{a}{2}, \frac{a}{2}\right)$, we have:

$$h(Z) = -\int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{1}{a}\ln\left(\frac{1}{a}\right) \ \mathrm{dz} = a\cdot\frac{1}{a}\cdot\ln a = \ln a \ \text{ nats}$$

So we need to compute the entropy of Y:

1. If $a = 1$:

$$\begin{aligned}
h(Y) &= -\int_{-1}^{0}(y+1)\ln(y+1) \ \mathrm{dy} - \int_0^1 (1-y)\ln(1-y) \ \mathrm{dy} \\
&= -\int_0^1 x\ln x \ \mathrm{dx} - \int_1^0 t\ln t(-1) \ \mathrm{dt} \qquad (\text{Let } x = y+1, t = 1-y) \\
&= -2I \\
&= \frac{1}{2} \ \text{ nat}
\end{aligned}$$

2. If $0 < a < 1$:

$$\begin{aligned}
h(Y) = &-\int_{-\frac{1}{2}-\frac{a}{2}}^{-\frac{1}{2}+\frac{a}{2}} \frac{1}{a}\left(y+\frac{1}{2}+\frac{a}{2}\right)\ln\left(\frac{1}{a}\left(y+\frac{1}{2}+\frac{a}{2}\right)\right) \ \mathrm{dy} - \int_{-\frac{1}{2}+\frac{a}{2}}^{\frac{1}{2}-\frac{a}{2}} 1\ln 1 \ \mathrm{dy} \\
&- \int_{\frac{1}{2}-\frac{a}{2}}^{\frac{1}{2}+\frac{a}{2}} \left(-\frac{1}{a}\left(y-\frac{1}{2}-\frac{a}{2}\right)\right)\ln\left(-\frac{1}{a}\left(y-\frac{1}{2}-\frac{a}{2}\right)\right) \ \mathrm{dy} \\
= &-a\int_0^1 x\ln x \ \mathrm{dx} + a\int_1^0 t\ln t \ \mathrm{dt} \qquad \left(\text{Let } x = \frac{1}{a}\left(y+\frac{1}{2}+\frac{a}{2}\right), t = -\frac{1}{a}\left(y-\frac{1}{2}-\frac{a}{2}\right)\right) \\
= &-a\cdot 2I \\
= &\frac{a}{2} \ \text{ nats}
\end{aligned}$$

3. If $a > 1$:

$$
\begin{aligned}
h(Y) = &-\int_{-\frac{1}{2}-\frac{a}{2}}^{\frac{1}{2}-\frac{a}{2}} \frac{1}{a}\left(y+\frac{1}{2}+\frac{a}{2}\right)\ln\left(\frac{1}{a}\left(y+\frac{1}{2}+\frac{a}{2}\right)\right)\ \mathrm{dy} - \int_{\frac{1}{2}-\frac{a}{2}}^{-\frac{1}{2}+\frac{a}{2}} \frac{1}{a}\ln\left(\frac{1}{a}\right)\ \mathrm{dy} \\
&- \int_{-\frac{1}{2}+\frac{a}{2}}^{\frac{1}{2}+\frac{a}{2}} \left(-\frac{1}{a}\left(y-\frac{1}{2}-\frac{a}{2}\right)\right)\ln\left(-\frac{1}{a}\left(y-\frac{1}{2}-\frac{a}{2}\right)\right)\ \mathrm{dy} \\
= &-\int_0^1 \left(\frac{1}{a}x\right)\ln\left(\frac{1}{a}x\right)\ \mathrm{dx} - \left(\frac{1}{a}\right)\ln\left(\frac{1}{a}\right)\cdot(a-1) - \int_1^0 \left(\frac{1}{a}t\right)\ln\left(\frac{1}{a}t\right)(-1)\ \mathrm{dt} \\
&\left(\text{Let } x = y+\frac{1}{2}+\frac{a}{2}, t = -\left(y-\frac{1}{2}-\frac{a}{2}\right)\right) \\
= &\,2\left(\frac{\ln a}{a}\int_0^1 x\ \mathrm{dx} - \frac{1}{a}\int_0^1 x\ln x\ \mathrm{dx}\right) + \frac{\ln a}{a}(a-1) \\
= &\ln a + \frac{1}{2a}\quad \text{nats}
\end{aligned}
$$

So above all, we have:

$$
I(X;Y) = \begin{cases} \frac{1}{2}\ \text{nat}, & a = 1 \\ \frac{a}{2} - \ln a\ \text{nats}, & 0 < a < 1 \\ \frac{1}{2a}\ \text{nats}, & a > 1 \end{cases}
$$

(b) Lemma: Among all distributions in the range $[a,b]$, the uniform distribution $\mathrm{Unif}(a,b)$ has the highest entropy.
Let $X_1 \sim \mathrm{Unif}(a,b)$ with PDF $u(x) = \dfrac{1}{b-a}, x \in [a,b]$, and X_2 be another random variable with PDF $f(x)$ in the same range. We have:

$$
h(X_1) = -\int_a^b u(x)\ln u(x)\ \mathrm{dx} = \ln(b-a)\quad \text{nats}
$$

Then we have:

$$
\begin{aligned}
0 &\le D\left(f\|u\right) \\
&= \int_a^b f(x)\ln\frac{f(x)}{u(x)}\ \mathrm{dx} \\
&= \int_a^b f(x)\ln f(x)\ \mathrm{dx} + \int_a^b f(x)\ln u(x)\ \mathrm{dx} \\
&= \ln(b-a) - h(X_2) \\
&= h(X_1) - h(X_2)
\end{aligned}
$$

So we have proved that $h(X_1) \ge h(X_2)$, if and only if $f(x) = u(x)$.
Since when $a = 1$, $I(X;Y) = h(Y) - \ln a = h(Y)$. To reach the capacity, i.e. the highest mutual information, we need to maximize the entropy of Y, i.e. construct $p(x)$ to make $Y \sim \mathrm{Unif}\left(-1,1\right)$.
We can construct that

$$
X = \begin{cases} -\frac{1}{2}, & \text{w.p. } \frac{1}{2} \\ \frac{1}{2}, & \text{w.p. } \frac{1}{2} \end{cases}
$$

Then we have:

$$
f_Y(y) = \sum_x f_{Y|X}(y|x)p(X=x) = \sum_z f_Z(y-x)p(X=x)
$$

When $y \in [-1, 0]$:

$$\left.\begin{array}{l} \text{if } X = -\dfrac{1}{2} \Rightarrow y - x \in \left[-\frac{1}{2}, \frac{1}{2}\right] \Rightarrow f_Z(y-x) = 1 \\ \text{if } X = \dfrac{1}{2} \Rightarrow y - x \in \left[-\frac{3}{2}, -\frac{1}{2}\right] \Rightarrow f_Z(y-x) = 0 \end{array}\right\} \Rightarrow f_Y(y) = \frac{1}{2}$$

And when $y \in [0, 1]$:

$$\left.\begin{array}{l} \text{if } X = -\dfrac{1}{2} \Rightarrow y - x \in \left[\frac{1}{2}, \frac{3}{2}\right] \Rightarrow f_Z(y-x) = 0 \\ \text{if } X = \dfrac{1}{2} \Rightarrow y - x \in \left[-\frac{1}{2}, \frac{1}{2}\right] \Rightarrow f_Z(y-x) = 1 \end{array}\right\} \Rightarrow f_Y(y) = \frac{1}{2}$$

So above all, we have proved that in such construction, $Y \sim \text{Unif}\,(-1, 1)$, and $C = \max_{p(x)} I(X;Y) = 1$ bit.

(c) Let $F(x)$ be the CDF of X, similarly to the discussion in the beginning, we have:

1. If $0 < a < 1$:

$$f_Y(y) = \begin{cases} \frac{1}{a} F\left(y + \frac{a}{2}\right), & -\frac{1}{2} - \frac{a}{2} \le y < -\frac{1}{2} + \frac{a}{2} \\ \frac{1}{a}\left[F\left(y + \frac{a}{2}\right) - F\left(y - \frac{a}{2}\right)\right], & -\frac{1}{2} + \frac{a}{2} \le y < \frac{1}{2} - \frac{a}{2} \\ \frac{1}{a}\left[1 - F\left(y - \frac{a}{2}\right)\right], & \frac{1}{2} - \frac{a}{2} \le y \le \frac{1}{2} + \frac{a}{2} \\ 0, & \text{otherwise} \end{cases}$$

2. If $a > 1$:

$$f_Y(y) = \begin{cases} \frac{1}{a} F\left(y + \frac{a}{2}\right), & -\frac{1}{2} - \frac{a}{2} \le y < \frac{1}{2} - \frac{a}{2} \\ \frac{1}{a}, & \frac{1}{2} - \frac{a}{2} \le y < -\frac{1}{2} + \frac{a}{2} \\ -\frac{1}{a}\left[1 - F\left(y - \frac{a}{2}\right)\right], & -\frac{1}{2} + \frac{a}{2} \le y \le \frac{1}{2} + \frac{a}{2} \\ 0, & \text{otherwise} \end{cases}$$

To make $Y \sim \text{Unif}\left(-\frac{a+1}{2}, \frac{a+1}{2}\right) \Rightarrow f_Y(y) = \dfrac{1}{a+1}$, we can put $f_Y(y) = \dfrac{1}{a+1}$ into the above equations and solve the equations to get the optimal $F(x)$.

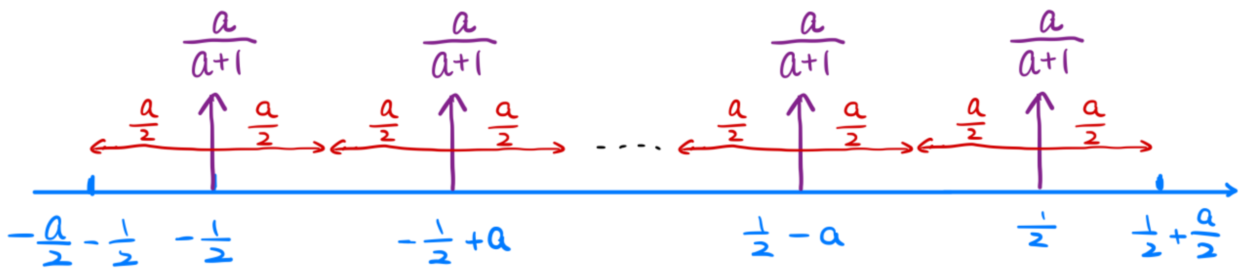
And we can get that:

1. When $0 < a < 1$:

$$F(x) = \begin{cases} 0, & x < -\frac{1}{2} \\ \dfrac{a}{a+1}, & -\frac{1}{2} \le x < -\frac{1}{2} + \frac{a}{2} \\ \vdots & \vdots \\ \dfrac{ia}{a+1}, & -\frac{1}{2} + i \cdot \frac{a}{2} \le x < -\frac{1}{2} + (i+1) \cdot \frac{a}{2} \\ \vdots & \vdots \\ 1, & \frac{1}{2} - \frac{a}{2} \le x \le \frac{1}{2} \end{cases}$$

Which could be intuitively understood as the following figure: We set $p(X = x) = \dfrac{a}{a+1}$, where $\mathcal{X}$ are the purple impulses signals in the figure. The red range represent that each $x \in \mathcal{X} = \{-\frac{1}{2}, -\frac{1}{2} + \frac{a}{2}, -\frac{1}{2} + \frac{3a}{2}, \ldots, -\frac{1}{2} + \frac{a}{2} + i \cdot a, \ldots, \frac{1}{2}\}$ could cover to make Y as uniform.

So in order to cover the full range in $[-\frac{a+1}{2}, \frac{a+1}{2}]$, we have to has $a = \dfrac{1}{N}$, where $N \in \{1, 2, \ldots\}$, and



the capacity is $C = \max_{p(x)} I(X;Y) = \log(a+1) = \log\left(\frac{1}{N}+1\right)$ bits. And in other situations, Y cannot be $\text{Unif}(-\frac{a+1}{2}, \frac{a+1}{2})$ because in such strategy, $\mathcal{X}$ could not cover the full range $\left[-\frac{a+1}{2}, \frac{a+1}{2}\right]$.

2. When $a > 1$:

Wwe could see that when $y \in \left[\frac{1}{2}-\frac{a}{2}, -\frac{1}{2}+\frac{a}{2}\right)$, $f_Y(y) = \dfrac{1}{a}$, unless $\frac{1}{2}-\frac{a}{2} = -\frac{1}{2}+\frac{a}{2} \Rightarrow a = 1$, it could be ignored, otherwise, Y would never be $\text{Unif}\left(-\frac{a+1}{2}, \frac{a+1}{2}\right)$.

So above all, we have proved that if and only if when $a = \dfrac{1}{N}, N \in \{1, 2, \ldots\}$, the capacity of the channel is $\log\left(\frac{1}{N}+1\right)$ bit.

And in other cases, we could not find a closed form solution for the $I(X;Y)$, so we need to use the Blahut-Arimoto algorithm mentioned in the text book §10.8.

Problem 3

8.8 Channel with uniformly distributed noise. Consider a additive channel whose input alphabet $\mathcal{X} = \{0, \pm1, \pm2\}$ and whose output $Y = X + Z$, where Z is distributed uniformly over the interval $[-1, 1]$. Thus, the input of the channel is a discrete random variable, whereas the output is continuous. Calculate the capacity $C = \max_{p(x)} I(X;Y)$ of this channel.

Solution

Since $Z \sim \text{Unif}[-1, 1]$, we have $f_Z(z) = \dfrac{1}{2}, z \in [-1, 1]$. So we have:

$$h(Z) = \int_{-1}^{1} f_Z(z) \log \frac{1}{f_Z(z)} dz = \log(1 - (-1)) = \log 2 = 1 \text{ bit}$$

So we can get that:

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(Z|X) \qquad \text{(since } Y = X + Z\text{)} \\ &= h(Y) - h(Z) \qquad \text{(since } Z \text{ is independent of } X\text{)} \\ &= h(Y) - 1 \end{aligned}$$

Let $p_i = p(X = i)$, then we can get that:

$$f_Y(y) = \sum_{x \in \mathcal{X}} f_Z(y - x) p(X = x) = \begin{cases} \frac{1}{2}p_{-2}, & -3 \le y \le -2 \\ \frac{1}{2}p_{-2} + \frac{1}{2}p_{-1}, & -2 < y \le -1 \\ \frac{1}{2}p_{-1} + \frac{1}{2}p_0, & -1 < y \le 0 \\ \frac{1}{2}p_0 + \frac{1}{2}p_1, & 0 < y \le 1 \\ \frac{1}{2}p_1 + \frac{1}{2}p_2, & 1 < y \le 2 \\ \frac{1}{2}p_2, & 2 < y \le 3 \end{cases}$$

As we have proved in Problem 2: if Y is a continous variable in range $[a, b]$, then $h(Y)_{\max} = \log(b - a)$. If and only if $Y \sim \text{Unif}(a, b)$.
So to reach the $h(Y)_{\max}$, we need to make $f_Y(y) = \dfrac{1}{6}$, $y \in [-3, 3]$.
So we have

$$p_{-2} = p_2 = \frac{1}{3}, p_{-1} = p_1 = 0, p_0 = \frac{1}{3}$$

So above all, we have proved that when $p(x) = \left(\dfrac{1}{3}, 0, \dfrac{1}{3}, 0, \dfrac{1}{3}\right)$

$$C = \max_{p(x)} I(X;Y) = h(Y) - 1 = \log 3 \text{ bits}$$

Problem 4

8.9 Gaussian mutual information. Suppose that (X, Y, Z) are jointly Gaussian and that $X \to Y \to Z$ forms a Markov chain. Let X and Y have correlation coefficient ρ_1 and let Y and Z have correlation coefficient ρ_2. Find $I(X; Z)$.

Solution

From what we have learned in the probability and statistics course, we know that the Minimum Mean Square Error Estimator (MMSE) of X given Y is $\mathbb{E}(X|Y)$. And the Linear Square Estimator (LLSE) of X given Y is $L(X|Y) = \mathbb{E}(Y) + \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}(Y - \mathbb{E}(Y))$.
Also, when X, Y are the jointly Gaussian random variables, the MMSE of X given Y and the LLSE of X given Y are the same. i.e.

$$\mathbb{E}(X|Y) = L(X|Y) = \mathbb{E}(Y) + \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}(Y - \mathbb{E}(Y))$$

Suppose $X \sim \mathcal{N}(\mu_X, \sigma_X^2), Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2), Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$.
Since phase shifing does not the variance and the covariance of the Gaussian random variables, we can let $\hat{X} = X - \mu_X \sim \mathcal{N}(0, \sigma_X^2), \hat{Y} = Y - \mu_Y \sim \mathcal{N}(0, \sigma_Y^2), \hat{Z} = Z - \mu_Z \sim \mathcal{N}(0, \sigma_Z^2)$.
And since the covariance between a random variable and a constant is 0, then we have

$$\text{Cov}\left(\hat{X}, \hat{Y}\right) = \text{Cov}(X-\mu_X, Y-\mu_Y) = \text{Cov}(X,Y) - \text{Cov}(\mu_X, Y) - \text{Cov}(X, \mu_Y) + \text{Cov}(\mu_X, \mu_Y) = \text{Cov}(X,Y) = \rho_1\sigma_X\sigma_Y$$

Similarly, we can get that

$$\text{Cov}\left(\hat{X}, \hat{Z}\right) = \text{Cov}(X,Z) = \rho_{X,Z}\sigma_X\sigma_Z, \text{Cov}\left(\hat{Y}, \hat{Z}\right) = \text{Cov}(Y,Z) = \rho_2\sigma_Y\sigma_Z$$

So we have

$$\begin{aligned}
\mathbb{E}(\hat{X}|\hat{Y}) &= \mathbb{E}(\hat{X}) + \frac{\text{Cov}(\hat{X},\hat{Y})}{\text{Var}(\hat{Y})}\left(\hat{Y} - \mathbb{E}(\hat{Y})\right) = \frac{\rho_1\sigma_X}{\sigma_Y}\hat{Y} \\
\mathbb{E}(\hat{Z}|\hat{Y}) &= \frac{\rho_2\sigma_Z}{\sigma_Y}\hat{Y}
\end{aligned}$$

Then we can get $\mathbb{E}\left(\hat{X}\hat{Z}\right)$ by applying the Law of Iterated Expectations. And the Markov Chain $\hat{X} \to \hat{Y} \to \hat{Z}$ also holds. i.e. $\hat{X} \perp \hat{Z}|\hat{Y}$. So we have

$$\begin{aligned}
\mathbb{E}\left(\hat{X}\hat{Z}\right) &= \mathbb{E}\left(\mathbb{E}\left(\hat{X}\hat{Z}|\hat{Y}\right)\right) && \text{(Law of Iterated Expectations)} \\
&= \mathbb{E}\left(\mathbb{E}\left(\hat{X}|\hat{Y}\right)\mathbb{E}\left(\hat{Z}|\hat{Y}\right)\right) && \left(\text{since } \hat{X} \perp \hat{Z}|\hat{Y}\right) \\
&= \mathbb{E}\left(\frac{\rho_1\sigma_X}{\sigma_Y}\hat{Y} \cdot \frac{\rho_2\sigma_Z}{\sigma_Y}\hat{Y}\right) \\
&= \frac{\rho_1\rho_2\sigma_X\sigma_Z}{\sigma_Y^2}\mathbb{E}\left(\hat{Y}^2\right) \\
&= \frac{\rho_1\rho_2\sigma_X\sigma_Z}{\sigma_Y^2}\sigma_Y^2 && \left(\text{since } \mathbb{E}\left(\hat{Y}\right) = 0 \Rightarrow \mathbb{E}\left(\hat{Y}^2\right) = \text{Var}\left(\hat{Y}\right)\right) \\
&= \rho_1\rho_2\sigma_X\sigma_Z
\end{aligned}$$

So

$$\rho_{X,Z} = \frac{\mathbb{E}\left(\hat{X}\hat{Z}\right)}{\sigma_X\sigma_Z} = \frac{\rho_1\rho_2\sigma_X\sigma_Z}{\sigma_X\sigma_Z} = \rho_1\rho_2$$

And we have known that the entropy of a Gaussian distribution $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ is $h(X) = \dfrac{1}{2}\ln\left(2\pi e \sigma_X^2\right)$. And the entropy of n random variable's jointly Gaussian with covariance matrix $\boldsymbol{K}$ is $\dfrac{1}{2}\ln\left((2\pi e)^n |\boldsymbol{K}|\right)$. And since the covariance matrix of X and Z is that

$$\boldsymbol{K} = \begin{bmatrix} \sigma_X^2 & \rho_{X,Z}\sigma_X\sigma_Z \\ \rho_{X,Z}\sigma_X\sigma_Z & \sigma_Z^2 \end{bmatrix}$$

So we can get that

$$\begin{aligned}
I(X;Z) &= h(X) + h(Z) - h(X,Z) \\
&= \frac{1}{2}\ln\left(2\pi e\sigma_X^2\right) + \frac{1}{2}\ln\left(2\pi e\sigma_Z^2\right) - \frac{1}{2}\ln\left((2\pi e)^2|\boldsymbol{K}|\right) \\
&= \frac{1}{2}\ln\left(\frac{\sigma_X^2\sigma_Z^2}{\sigma_X^2\sigma_Z^2 - \rho_{X,Z}^2\sigma_X^2\sigma_Z^2}\right) \\
&= \frac{1}{2}\ln\left(\frac{1}{1-\rho_{X,Z}^2}\right) \\
&= -\frac{1}{2}\ln\left(1-\rho_1^2\rho_2^2\right)
\end{aligned}$$