Tutorial 4

1 Exercises on Differential Entropy

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

   (a) The uniform distribution, $f(x) = \frac{1}{b-a}$.

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

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

   (d) The sum of X_1 and X_2, where X_1 and X_2 are independent normal random variables with mean μ_i and variance σ_i^2, $i = 1, 2$.

Solution:

(a) Uniform Distribution

$$\begin{aligned} h(f) &= -\int_a^b \frac{1}{b-a} \ln \frac{1}{b-a} dx \\ &= \ln(b-a) \quad \text{nats} \\ &= \log(b-a) \quad \text{bits} \end{aligned}$$

(b) Exponential distribution.

$$\begin{aligned} h(f) &= -\int_0^\infty \lambda e^{-\lambda x} \ln \lambda e^{-\lambda x} dx \\ &= -\int_0^\infty \lambda e^{-\lambda x} [\ln \lambda - \lambda x dx \\ &= -\ln \lambda + 1 \quad \text{nats} \\ &= \log \frac{e}{\lambda} \quad \text{bits} \end{aligned}$$

(c) Laplace density.

$$\begin{aligned} h(f) &= -\int_{-\infty}^\infty \frac{1}{2}\lambda e^{-\lambda|x|} \ln \frac{1}{2}\lambda e^{-\lambda|x|} dx \\ &= -\int_{-\infty}^\infty \frac{1}{2}\lambda e^{-\lambda|x|} [\ln \frac{1}{2} + \ln \lambda - \lambda|x|] dx \\ &= -\ln \frac{1}{2} - \ln \lambda + 1 \\ &= \ln \frac{2e}{\lambda} \quad \text{nats} \\ &= \log \frac{2e}{\lambda} \quad \text{bits} \end{aligned}$$

(d) The sum of two normal distributions.

The sum of two normal random variables is also normal, so applying the result derived the class for the normal distribution, since $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

$$h(f) = \frac{1}{2} \log 2\pi e(\sigma_1^2 + \sigma_2^2) \quad \text{bits}.$$

Remark: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $h(X) = \frac{1}{2}\log(2\pi e \sigma^2)$. □

2. Consider X is a continuous random variable defined over interval $[a, b]$.

   (a) What is the maximum value of $h(X)$?

   (b) What is the corresponding distribution of X?

   Solution:

   Let $u(x) = \frac{1}{b-a}$ be the uniform probability density function over $[a, b]$, and and let $p(x)$ be the probability mass function for X. Then

$$\begin{aligned} D(p||u) &= \int_a^b p(x) \log \frac{p(x)}{u(x)} dx \\ &= \int_a^b p(x) \log p(x) dx - \int_a^b p(x) \log u(x) dx \\ &= -h(X) - \int_a^b p(x) \log \frac{1}{b-a} dx \\ &= \log(b-a) - h(X) \end{aligned}$$

   Since $D(p||u) \geq 0$,

$$h(X) \leq \log(b-a)$$

   where the equality holds when $p(x)$ is the uniform probability density function over $[a, b]$. □

3. Consider a additive channel whose input alphabet $\mathcal{X} = \{0, \pm 1, \pm 2\}$, and whose output $Y = X + Z$, where Z is uniformly distributed over the interval $[-1, 1]$. Thus the input of the channel is a discrete random variable, while the output is continuous. Calculate the capacity $C = \max_{p(x)} I(X;Y)$ of the channel.

   Solution:

   We can expand the mutual information

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(Z) \end{aligned}$$

   and $h(Z) = \log 2$, since $Z \sim U(-1, 1)$.

   The output Y is a sum of a discrete and a continuous random variable, and if the probability of X are $p_{-2}, p_{-1}, \cdots, p_2$, then the output distribution of Y has a uniform distribution with weight $\frac{p_{-2}}{2}$ for $-3 \leq Y \leq -1$ when $X = -2$, uniform with weight $\frac{p_{-1}}{2}$ for $-2 \leq Y \leq 0$ when $X = -1$, uniform with weight $\frac{p_0}{2}$ for $-1 \leq Y \leq 1$ when $X = 0$, uniform with weight $\frac{p_1}{2}$ for $0 \leq Y \leq 2$ when $X = 1$, and uniform with weight $\frac{p_2}{2}$ for $1 \leq Y \leq 3$ when $X = 2$. Thus we have the density function of Y as follows

$$p_Y(y) = \begin{cases} \frac{p_{-2}}{2} & y \in [-3, -2) \\ \frac{p_{-2}+p_{-1}}{2} & y \in [-2, -1) \\ \frac{p_{-1}+p_0}{2} & y \in [-1, 0) \\ \frac{p_0+p_1}{2} & y \in [0, 1) \\ \frac{p_1+p_2}{2} & y \in [1, 2) \\ \frac{p_2}{2} & y \in [2, 3) \end{cases}$$

Given that Y ranges from $[-3, 3]$, the maximum entropy that it can have is an uniform over this range. This can be achieved if the distribution of X is $(1/3, 0, 1/3, 0, 1/3)$. Then $h(Y) = \log 6$ and the capacity of this channel is $C = \log 6 - \log 2 = \log 3$ bits. □

4. 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_{xy} and let Y and Z have correlation coefficient ρ_{yz}. Find I(X;Z).

   Solution:

   Note that for a constant a, $h(a+X) = h(X)$. Thus, without loss of generality, we assume that the means of X, Y and Z are zero. Let

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_x^2 & \sigma_x\sigma_z\rho_{xz} \\ \sigma_x\sigma_z\rho_{xz} & \sigma_z^2 \end{pmatrix}$$

   be the covariance matrix of X and Z where ρ_{xz} is the correlation coefficient between X and Z. Then we have

$$I(X; Z) = h(X) + h(Z) - h(X, Z)$$

   Since (X, Y, Z) are jointly Gaussian, X and Z are individually marginally Gaussian, and (X, Z) is jointly Gaussian. Thus, we have

$$\begin{aligned} I(X; Z) &= h(X) + h(Z) - h(X, Z) \\ &= \frac{1}{2}\log(2\pi e\sigma_x^2) + \frac{1}{2}\log(2\pi e\sigma_z^2) - \frac{1}{2}\log(2\pi e|\mathbf{\Lambda}|) \\ &= -\frac{1}{2}\log(1 - \rho_{xz}^2) \end{aligned}$$

   Now,

$$\begin{aligned} \rho_{xz} =& \frac{E[XZ]}{\sigma_x\sigma_z} \\ =& \frac{E[E[XZ|Y]]}{\sigma_x\sigma_z} \\ =& \frac{E[E[X|Y]E[Z|Y]]}{\sigma_x\sigma_z} \\ =& \frac{E[\frac{\sigma_x\rho_{xy}}{\sigma_y}Y]E[\frac{\sigma_z\rho_{yz}}{\sigma_y}Y]}{\sigma_x\sigma_z} \\ =& \frac{E[\frac{\sigma_x\sigma_z\rho_{xy}\rho_{yz}}{\sigma_y^2}Y^2]}{\sigma_x\sigma_z} \\ =& \frac{\frac{\sigma_x\sigma_z\rho_{xy}\rho_{yz}}{\sigma_y^2}E[Y^2]}{\sigma_x\sigma_z} \\ =& \frac{\frac{\sigma_x\sigma_z\rho_{xy}\rho_{yz}}{\sigma_y^2}\mathrm{Var}(Y)}{\sigma_x\sigma_z} \\ =& \rho_{xy}\rho_{yz} \end{aligned}$$

   We can conclude that

$$I(X; Y) = -\frac{1}{2}\log(1 - \rho_{xy}^2\rho_{yz}^2)$$

Remark: If (X, Y) is jointly Gaussian, the conditional distribution of X given $Y = y$ is as follows.

$$X|Y = y \sim \mathcal{N}\left(\mu_x + \frac{\sigma_x}{\sigma_y}\rho_{xy}(y - \mu_y), (1 - \rho_{xy}^2)\sigma_x^2\right)$$

□

2 Exercises on Gaussian Channel

1. Let Y_1 and Y_2 be conditionally independent and conditionally identically distributed given X.

   (a) Show $I(X; Y_1, Y_2) = 2I(X; Y_1) - I(Y_1; Y_2)$.

   (b) Conclude that the capacity of the channel $X \longrightarrow (Y_1, Y_2)$ is less than twice the capacity of the channel $X \longrightarrow Y_1$.

   Solution:

   (a)

$$\begin{aligned}
I(X; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X) \\
&= H(Y_1) + H(Y_2) - I(Y_1; Y_2) - (H(Y_1|X) + H(Y_2|X, Y_1)) \\
&= H(Y_1) + H(Y_2) - I(Y_1; Y_2) - H(Y_1|X) - H(Y_2|X) \\
&= H(Y_1) - H(Y_1|X) + H(Y_2) - H(Y_2|X) - I(Y_1; Y_2) \\
&= I(X; Y_1) + I(X; Y_2) - I(Y_1; Y_2) \\
&= 2I(X; Y_1) - I(Y_1; Y_2)
\end{aligned}$$

   (b) The capacity of the single look channel $X \longrightarrow Y_1$ is

$$C_1 = \max_{p(x)} I(X; Y_1)$$

   The capacity of the channel $X \longrightarrow (Y_1, Y_2)$ is

$$\begin{aligned}
C_2 &= \max_{p(x)} I(X; Y_1, Y_2) \\
&= \max_{p(x)} 2I(X; Y_1) - I(Y_1; Y_2) \\
&\leq \max_{p(x)} 2I(X; Y_1) \\
&= 2C_1
\end{aligned}$$

   Hence, the two independent looks cannot be more than twice as good as one look. □

2. Consider the ordinary Gaussian channel with two correlated looks at X, i.e., $Y = (Y_1, Y_2)$, where

$$\begin{aligned}
Y_1 &= X + Z_1 \\
Y_2 &= X + Z_2
\end{aligned}$$

with a power constraint P on X, and $(Z_1, Z_2) \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{K})$, where

$$\boldsymbol{K} = \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}.$$

Find the capacity C for

(a) $\rho = 1$

(b) $\rho = 0$

(c) $\rho = -1$

Solution:

It is clear that the input distribution that maximizes the capacity is $X \sim \mathcal{N}(0, P)$. Evaluating the mutual information for the distribution,

$$\begin{aligned} C &= \max I(X; Y_1, Y_2) \\ &= h(Y_1, Y_2) - h(Y_1, Y_2|X) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2|X) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2) \end{aligned}$$

Now since

$$(Z_1, Z_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}\right),$$

we have

$$h(Z_1, Z_2) = \frac{1}{2}\log(2\pi e)^2|\boldsymbol{K}| = \frac{1}{2}\log(2\pi e)^2 N^2(1-\rho^2).$$

Since $Y_1 = X + Z_1$ and $Y_2 = X + Z_2$, we have

$$(Y_1, Y_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} P+N & P+N\rho \\ P+N\rho & P+N \end{bmatrix}\right),$$

and

$$h(Y_1, Y_2) = \frac{1}{2}\log(2\pi e)^2|\boldsymbol{K}| = \frac{1}{2}\log(2\pi e)^2(N^2(1-\rho^2) + 2PN(1-\rho)).$$

Hence the capacity is

$$\begin{aligned} C &= h(Y_1, Y_2) - h(Z_1, Z_2) \\ &= \frac{1}{2}\log\left(1 + \frac{2P}{N(1+\rho)}\right). \end{aligned}$$

(a) $\rho = 1$. In this case, $C = \frac{1}{2}\log(1 + \frac{P}{N})$, which is the capacity of a single look channel. This is not surprising, since in this case $Y_1 = Y_2$.

(b) $\rho = 0$. In this case,

$$C = \frac{1}{2}\log\left(1 + \frac{2P}{N}\right),$$

which corresponds to using twice the power in a single look. The capacity is the same as the capacity of the channel $X \longrightarrow (Y_1 + Y_2)$.

(c) $\rho = 0$. In this case, $C = \infty$, which is not surprising since if we add Y_1 and Y_2, we can recover X exactly.

Remark: The capacity of the above channel in all cases is the same as the capacity f the channel $X \longrightarrow (Y_1 + Y_2)$. □

3. **Output power constraint.** Consider an additive white Gaussian noise channel with an expected output power constraint P. Thus $Y = X + Z$, $Z \sim \mathcal{N}(0, N)$, Z is independent of X, and $E[Y^2] \leq P$. Find the channel capacity.

Solution:

$$
\begin{aligned}
C &= \max_{p(X):E[(X+Z)^2]\leq P} I(X;Y) \\
&= \max_{p(X):E[(X+Z)^2]\leq P} h(Y) - h(Y|X) \\
&= \max_{p(X):E[(X+Z)^2]\leq P} h(Y) - h(Z|X) \\
&= \max_{p(X):E[(X+Z)^2]\leq P} h(Y) - h(Z)
\end{aligned}
$$

Given a constraint on the output power of Y, the maximum differential entropy is achieved by a normal distribution, and we can achieve this by have $X \sim \mathcal{N}(0, P - N)$, and in this case,

$$C = \frac{1}{2}\log 2\pi eP - \frac{1}{2}\log 2\pi eN = \frac{1}{2}\log\frac{P}{N}.$$

□

4. **Fading Channel.** Consider an additive fading channel

$$Y = XV + Z,$$

where Z is additive noise, V is a random variable representing fading, and Z and V are independent of each other and of X. Argue that knowledge of the fading factor V improves capacity by showing

$$I(X;Y|V) \geq I(X;Y).$$

Solution:

Expanding $I(X;Y,V)$ in two ways, we get

$$
\begin{aligned}
I(X;Y,V) =& I(X;V) + I(X;Y|V) \\
=& I(X;Y) + I(X;V|Y)
\end{aligned}
$$

i.e.

$$
\begin{aligned}
I(X;V) + I(X;Y|V) =& I(X;Y) + I(X;V|Y) \\
I(X;Y|V) =& I(X;Y) + I(X;V|Y) \\
I(X;Y|V) \geq& I(X;Y)
\end{aligned}
$$

□

5. Consider the additive whiter Gaussian channel $Y_i = X_i + Z_i$ where $Z_i \sim \mathcal{N}(0, N)$, and the input signal has average power constraint P.

   (a) Suppose we use all power at time 1, i.e. $E[X_1^2] = nP$ and $E[X_i^2] = 0$ for $i = 2, 3, \cdots, n$. Find

   $$\max_{p(x^n)} \frac{I(X^n; Y^n)}{n}$$

   where the maximization is over all distributions $p(x^n)$ subject to the constraint $E[X_1^2] = nP$ and $E[X_i^2] = 0$ for $i = 2, 3, \cdots, n$.

   (b) Find

   $$\max_{E[\frac{1}{n}\sum_i^n X_i^2] \leq P} \frac{I(X^n; Y^n)}{n}$$

   and compare to part (a).

Solution:

(a)

$$\begin{aligned}\max_{p(x^n)} \frac{I(X^n; Y^n)}{n} &= \max_{p(x^n)} \frac{I(X_1; Y_1)}{n} \\ &= \frac{\frac{1}{2}\log\left(1 + \frac{nP}{N}\right)}{n}\end{aligned}$$

where the first equality comes from the constraint that all our power, nP, be used at time 1, and the second equality comes from that fact that given Gaussian noise and a power constraint nP, $I(X; Y) \leq \frac{1}{2}\log(1 + \frac{nP}{N})$.

(b)

$$\begin{aligned}\max_{p(x^n)} \frac{I(X^n; Y^n)}{n} &= \max_{p(x^n)} \frac{nI(X; Y)}{n} \\ &= \max_{p(x^n)} I(X; Y) \\ &= \frac{1}{2}\log\left(1 + \frac{P}{N}\right)\end{aligned}$$

where the first equality comes from the fact that the channel is memoryless. Notice that the quantity in part (a) goes to 0 as $n \to \infty$ while the quantity in part (b) stays constant.

Remark: The impulse scheme is suboptimal. □