# CS258 Information Theory Homework 8

Zhou Litao 518030910407 F1803016

April 23, 2020

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

1. The exponential density, $f(x) = \lambda e^{-\lambda x}, x \geq 0$
2. The Laplace density, $f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$
3. The sum of $X_1$ and $X_2$, where $X_1$ and $X_2$ are independent normal random variables with means $\mu_i$ and variances $\sigma_i^2, i = 1, 2$

*Solution.*

1.
$$\begin{aligned}
h(X) &= -\int_0^\infty \lambda e^{-\lambda x} \ln \lambda e^{-\lambda x} dx \\
&= -\int_0^\infty \lambda e^{-\lambda x} (\ln \lambda - \lambda x) dx \\
&= \ln \lambda \int_0^\infty e^{-\lambda x} d(-\lambda x) - \lambda \int_0^\infty x d e^{-\lambda x} \\
&= \ln \lambda e^{-\lambda x}\Big|_0^\infty - \lambda x e^{-\lambda x}\Big|_0^\infty - \int_0^\infty e^{-\lambda x} d(-\lambda x) \\
&= -\ln \lambda + 1 = \ln \frac{e}{\lambda}
\end{aligned} \tag{1}$$

2.
$$\begin{aligned}
h(X) &= -\int_{-\infty}^{+\infty} \frac{1}{2}\lambda e^{-\lambda|x|} \ln \frac{1}{2}\lambda e^{-\lambda|x|} dx \\
&= -2\int_0^{+\infty} \frac{1}{2}\lambda e^{-\lambda x} \ln \frac{1}{2}\lambda e^{-\lambda x} dx \\
&= -\int_0^{+\infty} \lambda e^{-\lambda x} \ln \lambda e^{-\lambda x} dx + \ln 2 \int_0^\infty \lambda e^{-\lambda x} dx \\
&= \ln \frac{e}{\lambda} - \ln 2 \cdot e^{-\lambda x}\Big|_0^\infty \\
&= \ln \frac{2e}{\lambda}
\end{aligned} \tag{2}$$

3. By condition we know that
$$X_1, X_2 \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right) \tag{3}$$

Since $X_1 + X_2 = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, it follows that $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{(x-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}} \\
h(X) &= -\int f(x) \log f(x) dx \qquad (4) \\
&= -\int f(x) \log \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} + f(x) \left( -\frac{(x - (\mu_1 + \mu_2))^2}{2(\sigma_1^2 + \sigma_2^2)} \right) dx
\end{aligned}
$$

By property of normal distribution, we have

$$
\int f(x) dx = 1 \text{ and } \int (x - (\mu_1 + \mu_2))^2 f(x) dx = \sigma_1^2 + \sigma_2^2 \qquad (5)
$$

Hence,

$$
h(X) = \frac{1}{2} \log 2\pi(\sigma_1^2 + \sigma_2^2) + \frac{1}{2} \qquad (6)
$$

□

**Exercise 2** (Concavity of determinants) Let $K_1$ and $K_2$ be two symmetric nonnegative definite $n \times n$ matrices. Prove:

$$
\left|\lambda K_1 + \bar{\lambda} K_2\right| \geq |K_1|^{\lambda} |K_2|^{\bar{\lambda}} \quad \text{for } 0 \leq \lambda \leq 1, \quad \bar{\lambda} = 1 - \lambda
$$

where $|K|$ denotes the determinant of $K$. [Hint: Let $\mathbf{Z} = \mathbf{X}_\theta$ where $\mathbf{X}_1 \sim N(0, K_1), \mathbf{X}_2 \sim N(0, K_2)$ and $\theta =$ Bernoulli $(\lambda)$. Then use $h(\mathbf{Z}|\theta) \leq h(\mathbf{Z})$.]

*Proof.* Let $Z = \theta X_1 + (1 - \theta) X_2$, where $\theta, X_1$ and $X_2$ are independent, $\mathbf{X}_1 \sim N(0, K_1), \mathbf{X}_2 \sim N(0, K_2)$ and $\theta =$ Bernoulli $(\lambda)$. Note that for every entry in the covariance matrix of $Z$,

$$
\begin{aligned}
cov(Z_i, Z_j) &= E(Z_i Z_j) E((\theta X_{1i} + \bar{\theta} X_{2i})(\theta X_{1j} + \bar{\theta} X_{2j})) \\
&= E(\theta^2) E X_{1i} X_{1j} + E(\bar{\theta}^2) E X_{2i} X_{2j} \qquad (7) \\
&= \lambda cov(X_{1i}, X_{1j}) + \bar{\lambda} cov(X_{2i}, X_{2j})
\end{aligned}
$$

Hence we have $K_Z = \lambda K_{X_1} + \bar{\lambda} K_{X_2}$. Also note that $EZ = \mathbf{0}$. By Theorem 8.6.5 [Cover] we have that

$$
h(Z) \leq \frac{1}{2} \log(2\pi e)^n |K_Z| = \frac{1}{2} \log(2\pi e)^n \left|\lambda K_1 + \bar{\lambda} K_2\right| \qquad (8)
$$

By the formula for entropy of a multivariate normal distribution, we have that

$$
\begin{aligned}
h(X_1) &\leq \frac{1}{2} \log(2\pi e)^n |K_1| \\
h(X_2) &\leq \frac{1}{2} \log(2\pi e)^n |K_2| \\
\Rightarrow h(Z|\theta) &= \lambda h(X_1) + \bar{\lambda} h(X_2) \qquad (9) \\
&= \frac{\lambda}{2} \log(2\pi e)^n |K_1| + \frac{\bar{\lambda}}{2} \log(2\pi e)^n |K_2| \\
&= \frac{1}{2} \log(2\pi e)^n |K_1|^{\lambda} |K_2|^{\bar{\lambda}}
\end{aligned}
$$

Note that $h(Z|\theta) \leq h(Z)$. The result follows from Equation 8 and 9. □

**Exercise 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 $-a/2 \leq z \leq +a/2$

1. Find $I(X;Y)$ as a function of $a$
2. 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 $I(X;Y)$?
3. (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.*

1. We first calculate the distribution of $Y$.

$$p_Y(y) = \int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{1}{a} \mathbf{1}_{\{y-\frac{1}{2} \leq z \leq y+\frac{1}{2}\}} dz \tag{10}$$

If $a \leq 1$, we have that

$$p_Y(y) = \begin{cases} \frac{1}{a}\left(y + \frac{a+1}{2}\right) & -\frac{a+1}{2} \leq y \leq \frac{a-1}{2} \\ 1 & \frac{a-1}{2} < y \leq \frac{1-a}{2} \\ \frac{1}{a}\left(-y + \frac{a+1}{2}\right) & \frac{1-a}{2} < y \leq \frac{a+1}{2} \end{cases} \tag{11}$$

If $a > 1$, we have that

$$p_Y(y) = \begin{cases} \frac{1}{a}\left(y + \frac{a+1}{2}\right) & -\frac{a+1}{2} \leq y \leq \frac{1-a}{2} \\ \frac{1}{a} & \frac{1-a}{2} < y \leq \frac{a-1}{2} \\ \frac{1}{a}\left(-y + \frac{a+1}{2}\right) & \frac{a-1}{2} < y \leq \frac{a+1}{2} \end{cases} \tag{12}$$

For $a \leq 1$,

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) = h(Y) - h(Z) = h(Y) - \ln a \\ &= \int_{-\frac{a+1}{2}}^{\frac{a+1}{2}} p(y) \ln p(y) dy - \ln a \\ &= -2\int_0^{\frac{1-a}{2}} 1 \ln 1 dy - 2\int_{\frac{1-a}{2}}^{\frac{1+a}{2}} \frac{1}{a}(-y + \frac{a+1}{2}) \log \frac{1}{a}(-y + \frac{a+1}{2}) dy - \ln a \\ &= 0 - 2a \int_0^1 t \ln t dt - \ln a \quad (t \triangleq \frac{1}{a}\left(-y + \frac{a+1}{2})\right) \\ &= -2a\left(\frac{1}{2}t^2 \ln t - \frac{1}{4}t^2\right)\Big|_0^1 - \ln a \\ &= \frac{a}{2} - \ln a \end{aligned} \tag{13}$$

For $a > 1$,

$$\begin{aligned}
I(X;Y) &= h(Y) - h(Y|X) = h(Y) - h(Z) = h(Y) - \ln a \\
&= \int_{-\frac{a+1}{2}}^{\frac{a+1}{2}} p(y) \ln p(y) dy - \ln a \\
&= -2 \int_0^{\frac{a-1}{2}} \frac{1}{a} \ln \frac{1}{a} dy - 2 \int_{\frac{a-1}{2}}^{\frac{1+a}{2}} \frac{1}{a}(-y + \frac{a+1}{2}) \log \frac{1}{a}(-y + \frac{a+1}{2}) dy - \ln a \\
&= \frac{a-1}{a} \ln a - 2a \int_0^{\frac{1}{a}} t \ln t dt - \ln a \quad (t \triangleq \frac{1}{a} \left( -y + \frac{a+1}{2}) \right) \\
&= -\frac{1}{a} \ln a - 2a \left( \frac{1}{2} t^2 \ln t - \frac{1}{4} t^2 \right) \Big|_0^{\frac{1}{a}} \\
&= \frac{1}{2a}
\end{aligned} \tag{14}$$

2. Since $X$ and $Z$ are both limited to $\left[-\frac{1}{2}, \frac{1}{2}\right]$, $Y$ is limited to $[-1, 1]$. By the Example 12.2.4 in [Cover], the maximal differential entropy $h(Y)$ is $\ln 2$, which can be obtained when $Y$ is uniformly distributed, and thus $p(X = -\frac{1}{2}) = p(X = \frac{1}{2}) = \frac{1}{2}$.

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(Z) = h(Y) - \ln 1 \leq \ln 2 \text{ nats} \tag{15}$$

The capacity is $\ln 2$ nats $= 1$ bit.

3. If $a = \frac{1}{k}, k \in \mathbf{N}$, then we can construct $X$ to be uniformly distributed on $\left\{-\frac{1}{2}, -\frac{1}{2} + \frac{1}{k}, \ldots, \frac{1}{2}\right\}$, with $Y$ uniformly distributed on $\left[-\frac{1}{2} - \frac{1}{2k}, -\frac{1}{2} + \frac{1}{2k}\right]$. In this case, the maximal mutual information can be obtained. $C = \log\left(1 + \frac{1}{k}\right)$.

$\square$

**Exercise 4** (Channel with uniformly distributed noise) Consider a additive channel whose input alphabet $\mathcal{X} = \{0, \pm 1, \pm 2\}$ 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.* Note that

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(Z) = h(Y) - \log 2 \tag{16}$$

Note that the support set of $Y$ is $[-3, 3]$.By the Example 12.2.4 in [Cover], the maximal differential entropy $h(Y)$ is $\log 6$, which can be obtained when $p(X = -2) = p(X = 0) = p(X = 2) = \frac{1}{3}$, and $Y$ is uniformly distributed. The maximal mutual information is $\log 3$. $\square$

**Lemma 1** (Conditional Expectation of Two Normal Random Variables)**.** *Let $X$ and $Y$ be jointly Gaussian with variances $\sigma_1^2$ ,$\sigma_2^2$ and correlation coefficient $\rho$. We have that $E(X|Y) = \frac{\sigma_1 \rho}{\sigma_2} Y$.*

*Proof.* We calculate the conditional probability density of $p(x|y)$

$$\begin{aligned}
p(x|y) &= \frac{p(x,y)}{p(y)} \\
&= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{x^2}{\sigma_1^2} - 2\rho\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right]\right\}}{\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\frac{y^2}{\sigma_2^2}\right\}} \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{x^2}{\sigma_1^2} - 2\rho\frac{xy}{\sigma_1\sigma_2} + \frac{\rho^2 y^2}{\sigma_2^2}\right] - \frac{1}{2}\frac{y^2}{\sigma_2^2} + \frac{1}{2}\frac{y^2}{\sigma_2^2}\right\} \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)\sigma_1^2} \left(x - \frac{\sigma_1 \rho y}{\sigma_2}\right)\right\}
\end{aligned} \tag{17}$$

With $Y$ given, the mean value is $\frac{\sigma_1 \rho}{\sigma_2} Y$ $\square$

**Exercise 5** (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 $\rho_1$ and let $Y$ and $Z$ have correlation coefficient $\rho_2$. Find $I(X; Z)$

*Solution.* By the formula of mutual information between correlated Gaussian random variables, we have

$$I(X; Z) = -\frac{1}{2} \log(1 - \rho_{xz}^2) \tag{18}$$

With the lemma above, we can derive $\rho_{xz}$ as follows.

$$\begin{aligned}
\rho_{xz} &= \frac{\mathrm{E}\{XZ\}}{\sigma_x \sigma_z} \\
&= \frac{\mathrm{E}\{\mathrm{E}\{XZ|Y\}\}}{\sigma_x \sigma_z} && \text{Nested Expectation} \\
&= \frac{\mathrm{E}\{\mathrm{E}\{X|Y\}\mathrm{E}\{Z|Y\}\}}{\sigma_x \sigma_z} && \text{Markov Chains} \\
&= \frac{\mathrm{E}\left\{\left(\frac{\sigma_x \rho_{xy}}{\sigma_y} Y\right)\left(\frac{\sigma_z \rho_{zx}}{\sigma_y} Y\right)\right\}}{\sigma_x \sigma_z} && \text{Apply the Lemma} \\
&= \rho_{xy} \rho_{zy}
\end{aligned} \tag{19}$$

Hence $I(X; Z) = -\frac{1}{2} \log(1 - \rho_{xy}^2 \rho_{zy}^2)$ $\square$