

Homework 2

Hanmo Chen 2020214276

September 29, 2020

• **Acknowledgments:**

For the Gamma function and its series extensions, I refer to the website(<https://www.wolframalpha.com/>). And for the proof of Karamata's Inequality, I refer to the wikipedia(https://en.wikipedia.org/wiki/Karamata's_inequality). Some matrix tricks used in derivation comes from the Matrix Cookbook(<http://matrixcookbook.com/>)

• **Collaborators:** None

- *I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.*

Hanmo Chen

2.1. First, we have Chain Rule for mutual information

$$\begin{aligned} I(X_1, X_2, \dots, X_n; Y) &= \sum_{i=1}^n I(X_i; Y \mid X_{i-1}, X_{i-2}, \dots, X_1) \\ I(X; Y_1, Y_2, \dots, Y_n) &= \sum_{i=1}^n I(X; Y_i \mid Y_{i-1}, Y_{i-2}, \dots, Y_1) \end{aligned} \quad (1)$$

Using this Chain Rule twice,

$$\begin{aligned} I(X_1, \dots, X_n; Y_1, \dots, Y_m) &= \sum_{j=1}^m I(X_1, X_2, \dots, X_n; Y_j \mid Y_{j-1}, Y_{j-2}, \dots, Y_1) \\ &= \sum_{i=1}^n \sum_{j=1}^m I(X_i; Y_j \mid X_{i-1}, X_{i-2}, \dots, X_1, Y_{j-1}, Y_{j-2}, \dots, Y_1) \end{aligned} \quad (2)$$

2.2. Examples:

(a) $I(X; Y \mid Z) < I(X; Y)$

Let $Z = X$ then $I(X; Y \mid Z) = H(X \mid Z) - H(X, Y \mid Z) = 0$. But for X, Y that are not independent, $I(X; Y) > 0 = I(X; Y \mid Z)$

- (b) $I(X; Y | Z) > I(X; Y)$

Use the example illustrated in class. let X and Y be independent Bernoulli random variables with $p = 0.5$, and let $Z = X + Y$. Then $I(X; Y) = 0$, but $I(X; Y | Z) = H(X | Z) - H(X | Y, Z) = H(X | Z) = 0.5$ bit.

2.3. Information Measures.

- (a) Denote $n = |A|$, because Z_1, Z_2, \dots, Z_n are i.i.d Bern($\frac{1}{2}$) random variables, so $X_A = (Z_i)_{i \in A}$ is a n -dimension random vector having 2^n possible points with equal probability,

$$H(X_A) = 2^n * (-\frac{1}{2^n} \log \frac{1}{2^n}) = n = |A| \text{ (bits)} \quad (3)$$

- (b) Denote $A = A_1 \cup A_2, A^* = A_1 \cap A_2, A_1^* = A_1 \setminus A_2, A_2^* = A_2 \setminus A_1$, so $X_{A_1} = (X_{A_1^*}, X_{A_1^*})$, $X_{A_2} = (X_{A_2^*}, X_{A_2^*})$. Because $X_{A_1^*}, X_{A_1^*}, X_{A_2^*}$ are mutually independent,

$$\begin{aligned} H(X_{A_1} | X_{A_2}) &= H((X_{A_1^*}, X_{A_1^*}) | (X_{A_2^*}, X_{A_2^*})) \\ &= H(X_{A_1^*}) = |A_1^*| = |A_1 \setminus A_2| \end{aligned} \quad (4)$$

Thus,

$$\begin{aligned} H(X_{A_1}, X_{A_2}) &= H(X_{A_2}) + H(X_{A_1} | X_{A_2}) \\ &= |A_2| + |A_1^*| = |A_1 \cup A_2| \end{aligned} \quad (5)$$

$$\begin{aligned} I(X_{A_1}; X_{A_2}) &= H(X_{A_2}) - H(X_{A_2} | X_{A_1}) \\ &= |A_2| - |A_2^*| = |A_1 \cap A_2| \end{aligned} \quad (6)$$

- 2.4. (a) No. $f_{X,Y}(x, y) = 0.5, |x| + |y| \leq 1$, and $f_X(x) = 1 - |x|, x \in [-1, 1], f_Y(y) = 1 - |y|, y \in [-1, 1]$, $f_{X,Y}(x, y) \neq f_X(x)f_Y(y)$ so X, Y are not independent.
- (b) Because X and Y are symmetric, $H(X) = H(Y)$ and $I(X; Y) = H(X) + H(Y) - H(X, Y) = 2H(X) - H(X, Y)$. Denote the area of B_p as $S_p = 4 \int_0^1 (1 - |x|^p)^{1/p} dx$, joint distribution of (X, Y) is $f_p(x, y) = \frac{1}{S_p}, |x|^p + |y|^p \leq 1$ the *pdf* of X is $f_p(x) = \frac{2}{S_p}(1 - |x|^p)^{1/p}, x \in [-1, 1]$
- So

$$I(X; Y) = 2H(X) - H(X, Y) = -2 \int_{-1}^1 f_p(x) \log(f_p(x)) dx - \log S_p \quad (7)$$

- $p = \frac{1}{2}, S_p = \frac{2}{3}, I(X; Y) = \frac{7}{3} - \ln 3 - \ln 2 = 0.542$
- $p = 1, S_p = 2, I(X; Y) = 1 - \ln 2 = 0.307$
- $p = \infty, X, Y$ are independent, $I(X; Y) = 0$

$$(c) \quad \lim_{p \rightarrow 0} I(X; Y) = \infty \quad (8)$$

Denote $n = 1/p$,

$$S_p = 4 \int_0^1 (1 - x^{\frac{1}{n}})^n dx = 4n \int_0^1 t^n (1 - t)^{n-1} dt = 4 \frac{(\Gamma(n+1))^2}{\Gamma(2n+1)} \quad (9)$$

And

$$I_n(X; Y) = 2n \left(\sum_{i=n+1}^{2n} \frac{1}{i} \right) - \log \frac{(2n)!}{(n!)^2} \quad (10)$$

Because

$$\lim_{n \rightarrow \infty} \sum_{i=n+1}^{2n} \frac{1}{i} = \ln 2 \quad (11)$$

$$\log \frac{(2n)!}{(n!)^2} = \log \left(\frac{2^{2n} \Gamma(n + \frac{1}{2})}{\Gamma(n+1)} \right) - \frac{\log(\pi)}{2} \quad (12)$$

And $\frac{\Gamma(n + \frac{1}{2})}{\Gamma(n+1)} \sim \log(n)$ as $n \rightarrow \infty$, thus

$$\lim_{n \rightarrow \infty} I_n(X; Y) = \lim_{n \rightarrow \infty} \frac{1}{2} \log(n) = \infty \quad (13)$$

2.5. The density function of Gaussian distribution is

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{m})^T \Sigma^{-1} (\mathbf{x} - \mathbf{m}) \right) \quad (14)$$

Thus,

$$\begin{aligned} D(\mathcal{N}(\mathbf{m}_1, \Sigma_1) \parallel \mathcal{N}(\mathbf{m}_0, \Sigma_0)) &= \mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \Sigma_1)} \left[\log \left(\frac{p_1(\mathbf{x})}{p_0(\mathbf{x})} \right) \right] \\ &= \mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \Sigma_1)} \left[\frac{1}{2} \log \frac{|\Sigma_0|}{|\Sigma_1|} - \frac{1}{2} (\mathbf{x} - \mathbf{m}_1)^T \Sigma_1^{-1} (\mathbf{x} - \mathbf{m}_1) \right. \\ &\quad \left. + \frac{1}{2} (\mathbf{x} - \mathbf{m}_0)^T \Sigma_0^{-1} (\mathbf{x} - \mathbf{m}_0) \right] \\ &= \frac{1}{2} \log \frac{|\Sigma_0|}{|\Sigma_1|} - \frac{1}{2} \mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \Sigma_1)} \left[(\mathbf{x} - \mathbf{m}_1)^T \Sigma_1^{-1} (\mathbf{x} - \mathbf{m}_1) \right] \\ &\quad + \frac{1}{2} \mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \Sigma_1)} \left[(\mathbf{x} - \mathbf{m}_0)^T \Sigma_0^{-1} (\mathbf{x} - \mathbf{m}_0) \right] \end{aligned} \quad (15)$$

For $\mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \Sigma_1)} \left[(\mathbf{x} - \mathbf{m}_1)^T \Sigma_1^{-1} (\mathbf{x} - \mathbf{m}_1) \right]$,

$$\begin{aligned}
\mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)} \left[(\mathbf{x} - \mathbf{m}_1)^T \mathbf{\Sigma}_1^{-1} (\mathbf{x} - \mathbf{m}_1) \right] &= \text{tr} \left(\mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)} \left[(\mathbf{x} - \mathbf{m}_1)^T \mathbf{\Sigma}_1^{-1} (\mathbf{x} - \mathbf{m}_1) \right] \right) \\
&= \mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)} \left[\text{tr} \left((\mathbf{x} - \mathbf{m}_1)^T \mathbf{\Sigma}_1^{-1} (\mathbf{x} - \mathbf{m}_1) \right) \right] \\
&= \mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)} \left[\text{tr} \left((\mathbf{x} - \mathbf{m}_1) (\mathbf{x} - \mathbf{m}_1)^T \mathbf{\Sigma}_1^{-1} \right) \right] \\
&= \text{tr} \left(\mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)} \left[(\mathbf{x} - \mathbf{m}_1) (\mathbf{x} - \mathbf{m}_1)^T \right] \mathbf{\Sigma}_1^{-1} \right) \\
&= \text{tr} (\mathbf{\Sigma}_1 \mathbf{\Sigma}_1^{-1}) = \text{tr}(\mathbf{I}_n) = n
\end{aligned} \tag{16}$$

In the same way,

$$\begin{aligned}
\mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)} \left[(\mathbf{x} - \mathbf{m}_0)^T \mathbf{\Sigma}_0^{-1} (\mathbf{x} - \mathbf{m}_0) \right] &= \text{tr} \left(\mathbb{E}_{\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)} \left[(\mathbf{x} - \mathbf{m}_0) (\mathbf{x} - \mathbf{m}_0)^T \right] \mathbf{\Sigma}_0^{-1} \right) \\
&= \text{tr} \left(\left((\mathbf{m}_1 - \mathbf{m}_0) (\mathbf{m}_1 - \mathbf{m}_0)^T + \mathbf{\Sigma}_1 \right) \mathbf{\Sigma}_0^{-1} \right) \\
&= \text{tr} (\mathbf{\Sigma}_1 \mathbf{\Sigma}_0^{-1}) + (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{\Sigma}_0^{-1} (\mathbf{m}_1 - \mathbf{m}_0)
\end{aligned} \tag{17}$$

So

$$D(\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1) \parallel \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)) = \frac{1}{2} \left[\log \frac{|\mathbf{\Sigma}_0|}{|\mathbf{\Sigma}_1|} - n + \text{tr} (\mathbf{\Sigma}_1 \mathbf{\Sigma}_0^{-1}) + (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{\Sigma}_0^{-1} (\mathbf{m}_1 - \mathbf{m}_0) \right] \tag{18}$$

(a) $D(\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1) \parallel \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)) < \infty$ when $\mathbf{\Sigma}_1, \mathbf{\Sigma}_0$ is non-singular.

(b) With $\mathbf{m}_0 = 0, \mathbf{\Sigma}_0 = \mathbf{I}_n$ in the equation [18],

$$D(\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}_n)) = \frac{1}{2} \left[-\log |\mathbf{\Sigma}_1| - n + \text{tr} (\mathbf{\Sigma}_1) + \mathbf{m}_1^T \mathbf{m}_1 \right] \tag{19}$$

(c) As the equation [18] shows,

$$D(\mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1) \parallel \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)) = \frac{1}{2} \left[\log \frac{|\mathbf{\Sigma}_0|}{|\mathbf{\Sigma}_1|} - n + \text{tr} (\mathbf{\Sigma}_1 \mathbf{\Sigma}_0^{-1}) + (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{\Sigma}_0^{-1} (\mathbf{m}_1 - \mathbf{m}_0) \right] \tag{20}$$

2.6. (a) Actually, it is called Karamata's Inequality, and the proof is given as below.

Define

$$c_i = \begin{cases} \frac{f(P_i) - f(Q_i)}{P_i - Q_i}, & \text{if } P_i \neq Q_i \\ f'(P_i), & \text{if } P_i = Q_i \end{cases} \tag{21}$$

Because f is a convex function, $\frac{f(x) - f(y)}{x - y}$ is monotonically non-decreasing for both x and y , we have $c_i \geq c_{i+1}$

Define $A_0 = B_0 = 0$, $A_n = \sum_{i=1}^n P_i, B_n = \sum_{i=1}^n Q_i$, so $A_n \leq B_n, 1 \leq n \leq k$, and $A_k = B_k$.

Consider

$$\begin{aligned}
\sum_{i=1}^k (f(P_i) - f(Q_i)) &= \sum_{i=1}^n c_i (P_i - Q_i) \\
&= \sum_{i=1}^n c_i (\underbrace{A_i - A_{i-1}}_{=P_i} - \underbrace{(B_i - B_{i-1})}_{=Q_i}) \\
&= \sum_{i=1}^n c_i (A_i - B_i) - \sum_{i=1}^n c_i (A_{i-1} - B_{i-1}) \\
&= c_n (\underbrace{A_n - B_n}_{=0}) + \sum_{i=1}^{n-1} (\underbrace{c_i - c_{i+1}}_{\geq 0}) (\underbrace{A_i - B_i}_{\leq 0}) - c_1 (\underbrace{A_0 - B_0}_{=0}) \\
&\leq 0
\end{aligned} \tag{22}$$

Therefore,

$$\sum_{i=1}^k f(P_i) \leq \sum_{i=1}^k f(Q_i) \tag{23}$$

(b) Let $f(x) = x \log x$ is convex, using 23 and $H(P) = -\sum_{i=1}^k f(P_i)$, so

$$H(P) \geq H(Q) \tag{24}$$