Tsinghua-Berkeley Shenzhen Institute
INFORMATION THEORY AND STATISTICAL LEARNING
Fall 2020

Homework 3

Chenyu Tian October 21, 2020

- **Acknowledgments:** This template takes some materials from course CSE 547/Stat 548 of Washington University:
`https://courses.cs.washington.edu/courses/cse547/17sp/index.html`.

- *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.*

3.1.

$$\begin{aligned}
&I\left(X_1,\ldots,X_n;Y_1,\ldots,Y_m\right)\\
=&\sum_{i=1}^{n} I\left(X_i;Y_1,\ldots,Y_m|X_{i-1},\ldots,X_1\right)\\
=&\sum_{i=1}^{n} I\left(Y_1,\ldots,Y_m;X_i|X_{i-1},\ldots,X_1\right)\\
=&\sum_{i=1}^{n}(H\left(Y_1,\ldots,Y_m|X_{i-1},\ldots,X_1\right)-H\left(Y_1,\ldots,Y_m|X_i,\ldots,X_1\right))\\
=&\sum_{i=1}^{n}(\sum_{j=1}^{m} H\left(Y_j|Y_{j-1},\ldots,Y_1|X_{i-1},\ldots,X_1\right)-\sum_{j=1}^{m} H\left(Y_j|Y_{j-1},\ldots,Y_1|X_i,\ldots,X_1\right))\\
=&\sum_{i=1}^{n}\sum_{j=1}^{m}(H\left(Y_j|Y_{j-1},\ldots,Y_1|X_{i-1},\ldots,X_1\right)-\sum_{j=1}^{m} H\left(Y_j|Y_{j-1},\ldots,Y_1|X_i,\ldots,X_1\right))\\
=&\sum_{i=1}^{n}\sum_{j=1}^{m} I\left(X_i;Y_j|X_1,\ldots,X_{i-1};Y_1,\ldots,Y_{j-1}\right)
\end{aligned} \tag{1}$$

3.2. (a) Define $X = \begin{cases} 0 & w.p. \frac{1}{2} \\ 1 & w.p. \frac{1}{2} \end{cases}$, $Y = 2X$ and $Z = 2Y$. Then $I(X;Y) = H(X) + H(Y) - H(X,Y) = 1$.

$$\begin{aligned}
I(X;Y|Z) &= H(X|Z) - H(X|Y,Z)\\
&= H(X,Z) - H(Z) - (H(X,Y,Z) + H(Y,Z))\\
&= 0
\end{aligned} \tag{2}$$

$I(X;Y|Z) < I(X;Y)$.

(b) Define $X = Y = \left\{ \begin{array}{ll} 0 & w.p.\frac{1}{2} \\ 1 & w.p.\frac{1}{2} \end{array} \right.$, $\quad Z = X + Y$ X and Y are independent. Then $I(X;Y) = 0$.

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &= H(X,Z) - H(Z) - (H(X,Y,Z) + H(Y,Z)) \\ &= 2 - \frac{2}{3} - 2 + 2 \\ &= \frac{1}{2} \end{aligned} \tag{3}$$

$I(X;Y|Z) > I(X;Y)$.

3.3. (a) Since $Z_1, \ldots, Z_n$ are i.i.d and $Bern(\frac{1}{2})$ distributed, we can get

$$\begin{aligned} H\left(Z_i\right) &= -\sum_{i=1}^{2} \frac{1}{2} \log_2 \left(\frac{1}{2}\right) = 1 \quad \forall i \in 1, 2, \ldots, n \\ \rightarrow H\left(X_A\right) &= H\left(\left(Z_i\right)_{i \in A}\right) \\ &= \sum_{i \in A} H\left(Z_i\right) \\ &= |A| \end{aligned} \tag{4}$$

(b)

$$\begin{aligned} H\left(X_{A_1}, X_{A_2}\right) &= H\left(\left(Z_i\right)_{i \in A_1}, \left(Z_i\right)_{i \in A_2}\right) \\ &= H\left(\left(Z_i\right)_{i \in A_1 \cup A_2}\right) \\ &= \left|A_1 \cup A_2\right| \\ H\left(X_{A_1} \mid X_{A_2}\right) &= H\left(X_{A_1}, X_{A_2}\right) - H\left(X_{A_2}\right) \\ &= \left|A_1 \cup A_2\right| - \left|A_2\right| \\ &= \left|A_1 \backslash A_2\right| \\ I\left(X_{A_1}; X_{A_2}\right) &= H\left(X_{A_1}\right) - H\left(X_{A_1} \mid X_{A_2}\right) \\ &= \left|A_1\right| - \left|A_1 \backslash A_2\right| \\ &= \left|A_1 \cap A_2\right| \end{aligned} \tag{5}$$

3.4. (a) When $p = 1$,

$$P(x,y) = \begin{array}{ll} \frac{1}{2} & |x| + |y| \leq 1 \\ 0 & else \end{array} . \tag{6}$$

Let $x = 0, y = 1$, we have $P(x,y) = \frac{1}{2}, P_X(0) = 1, P_Y(1) = 0$. Then $P(1,0) \neq P_X(0) * P_Y(1)$ and X and Y are not independent.

(b) When $p = \frac{1}{2}$,

$$\begin{aligned} H(X,Y) &= \iint_S \frac{3}{2} \log_2 \frac{2}{3} \mathrm{d}x\mathrm{d}y = \log_2 \frac{2}{3} \quad S : |x|^{\frac{1}{2}} + |y|^{\frac{1}{2}} \leq 1 \\ H(Y) = H(X) &= 2 * \int_0^1 \frac{3 * 2 * (1 - x^{\frac{1}{2}})^2}{2} \log_2 \frac{2}{3 * 2 * (1 - x^{\frac{1}{2}})^2} \mathrm{d}x = \end{aligned} \tag{7}$$

Thus, $I(X,Y) = H(X) + H(Y) - H(X,Y) =$.
When $p = 1$, as the same procedure,
$I(X,Y) = H(X) + H(Y) - H(X,Y) = 1 + 1 - 1 = 1$.
When $p = \infty$, as the same procedure,
$I(X,Y) = H(X) + H(Y) - H(X,Y) = 1 + 1 - 2 = 1$.

(c) $I(X;Y)$ converge to ∞.

3.5.

$$
\begin{aligned}
& D\left(\mathcal{N}\left(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1\right) \| \mathcal{N}\left(\boldsymbol{m}_0, \boldsymbol{\Sigma}_0\right)\right) \\
=& \int \left[\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}\left(x-\boldsymbol{m}_0\right)^T \boldsymbol{\Sigma}_0^{-1}\left(x-\boldsymbol{m}_0\right) + \frac{1}{2}\left(x-\boldsymbol{m}_1\right)^T \boldsymbol{\Sigma}_1^{-1}\left(x-\boldsymbol{m}_1\right)\right] \times p(x)dx \\
=& \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}\operatorname{tr}\left\{E\left[\left(x-\boldsymbol{m}_0\right)\left(x-\boldsymbol{m}_0\right)^T\right]\boldsymbol{\Sigma}_0^{-1}\right\} + \frac{1}{2}E\left[\left(x-\boldsymbol{m}_1\right)^T\boldsymbol{\Sigma}_1^{-1}\left(x-\boldsymbol{m}_1\right)\right] \\
=& \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}\operatorname{tr}\left\{I_n\right\} + \frac{1}{2}\left(\boldsymbol{m}_0-\boldsymbol{m}_1\right)^T\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{m}_0-\boldsymbol{m}_1\right) + \frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\right\} \\
=& \frac{1}{2}\left[\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - n + \operatorname{tr}\left\{\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\right\} + \left(\boldsymbol{m}_1-\boldsymbol{m}_0\right)^T\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{m}_1-\boldsymbol{m}_0\right)\right]
\end{aligned}
\tag{8}
$$

(a) When $\boldsymbol{\Sigma}_0$ is non-singular.

(b) Based on 8,
$D\left(\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma}) \| \mathcal{N}\left(0, \boldsymbol{I}_n\right)\right) = \frac{1}{2}\left[\log|\boldsymbol{\Sigma}| - n + \operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\right\} + \boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right]$.

(c) Based on 8, $D\left(\mathcal{N}\left(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1\right) \| \mathcal{N}\left(\boldsymbol{m}_0, \boldsymbol{\Sigma}_0\right)\right) =$
$\frac{1}{2}\left[\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - n + \operatorname{tr}\left\{\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\right\} + \left(\boldsymbol{m}_1-\boldsymbol{m}_0\right)^T\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{m}_1-\boldsymbol{m}_0\right)\right]$.

3.6. (a)

$$
\begin{aligned}
& \sum_{i=1}^{k} f\left(P_i\right) + \sum_{i=1}^{k} f\left(Q_i\right) \\
\geq & 2\sum_{i=1}^{k} f\left(\frac{P_i + Q_i}{2}\right)
\end{aligned}
\tag{9}
$$

$$
\Rightarrow \sum_{i=1}^{k} f\left(P_i\right) - \sum_{i=1}^{k} f\left(\frac{P_i+Q_i}{2}\right) \geq \sum_{i=1}^{k} f\left(\frac{P_i+Q_i}{2}\right) - \sum_{i=1}^{k} f\left(Q_i\right)
$$

Thus, $\sum_{i=1}^{k} f\left(P_i\right) \leq \sum_{i=1}^{k} f\left(Q_i\right)$.

(b) Define $f(x) = x\log x$, which is a convex function because $f''(x) = \frac{1}{x} \leq 0$. Based on (a), it has

$$
\begin{aligned}
& \sum_{i=1}^{k} f\left(P_i\right) \leq \sum_{i=1}^{k} f\left(Q_i\right) \\
\Rightarrow & -\sum_{i=1}^{k} P_i \log P_i \geq -\sum_{i=1}^{k} Q_i \log Q_i \\
\Rightarrow & H(P) \geq H(Q)
\end{aligned}
\tag{10}
$$

3.7. (a)

$$
\begin{aligned}
& C\left(X_1, X_2, \ldots, X_n\right) \\
= & D\left(P_{X^n} \| \prod_{i=1}^n P_{X_i}\right) \\
= & \sum_{x_1 \in X} \cdots \sum_{x_n \in X} P_{X^n}\left(x_1, \ldots, x_n\right) \log \frac{P_{X^n}\left(x_1, \ldots, x_n\right)}{\prod_{i=1}^n P_{X_i}\left(x_i\right)} \\
= & -\sum_{x_1 \in X} \cdots \sum_{x_n \in X} P_{X^n}\left(x_1, \ldots, x_n\right) \log \prod_{i=1}^n P_{X_i}\left(x_i\right) \\
& +\sum_{x_1 \in X} \cdots \sum_{x_n \in X} P_{X^n}\left(x_1, \ldots, x_n\right) \log P_{X^n}\left(x_1, \ldots, x_n\right) \\
= & -\sum_{i=1}^n \sum_{x_1 \in X} \cdots \sum_{x_n \in X} P_{X^n}\left(x_1, \ldots, x_n\right) \log P_{X_i}\left(x_i\right) \\
& +\sum_{x_1 \in X} \cdots \sum_{x_n \in X} P_{X^n}\left(x_1, \ldots, x_n\right) \log P_{X^n}\left(x_1, \ldots, x_n\right) \\
= & -\sum_{i=1}^n \sum_{x_i \in X} P_{X_i}\left(x_i\right) \log P_{X_i}\left(x_i\right) \\
& +\sum_{x_1 \in X} \cdots \sum_{x_n \in X} P_{X^n}\left(x_1, \ldots, x_n\right) \log P_{X^n}\left(x_1, \ldots, x_n\right) \\
= & \sum_{i=1}^n H\left(X_i\right)-H\left(X_1, \ldots, X_n\right) \\
= & \sum_{i=1}^n H\left(X_i\right)-\sum_{i=1}^n H\left(X_i \mid X_1, \ldots, X_{i-1}\right) \\
= & \sum_{i=1}^n\left(H\left(X_i\right)-H\left(X_i \mid X_1, \ldots, X_{i-1}\right)\right) \\
= & \sum_{i=1}^{n-1}\left(H\left(X_{i+1}\right)-H\left(X_{i+1} \mid X_1, \ldots, X_i\right)\right) \\
= & \sum_{i=1}^{n-1} I\left(X^i ; X_{i+1}\right)
\end{aligned} \tag{11}
$$

(b) When n random variables are independent.

3.8. (a)

(b)