ECE 5630 - Solutions Homework Assignment 3

1) a) Using the fact that $p_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \prod_{i=1}^k p_{X_i|X_1,\ldots,X_{i-1}}(x_i|x_1,\ldots,x_{i-1})$ for all $(x_1,\ldots,x_k) \in \mathcal{X}^k$, and the linearity of expectation, we get

$$\begin{aligned} H(X_1,\ldots,X_k) &= \mathbb{E}_{P_{X_1,\ldots,X_k}}\left[\log\left(\frac{1}{p_{X_1,\ldots,X_k}(X_1,\ldots,X_k)}\right)\right] \\ &= \mathbb{E}_{P_{X_1,\ldots,X_k}}\left[\log\left(\frac{1}{\prod_{i=1}^k p_{X_i|X_1,\ldots,X_{i-1}}(X_i|X_1,\ldots,X_{i-1})}\right)\right] \\ &= \sum_{i=1}^k \mathbb{E}_{P_{X_1,\ldots,X_k}}\left[\log\left(\frac{1}{p_{X_i|X_1,\ldots,X_{i-1}}(X_i|X_1,\ldots,X_{i-1})}\right)\right] \\ &= \sum_{i=1}^k \mathbb{E}_{P_{X_1,\ldots,X_i}}\left[\log\left(\frac{1}{p_{X_i|X_1,\ldots,X_{i-1}}(X_i|X_1,\ldots,X_{i-1})}\right)\right], \end{aligned}$$

where the last equality follows from the marginalization of the joint PMF, i.e.,

$$\sum_{(x_{i+1},\ldots,x_k)\in\mathcal{X}^{k-i}} p_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = p_{X_1,\ldots,X_i}(x_1,\ldots,x_i).$$

Thus, using the definition of conditional entropy we get

$$H(X_1,\ldots,X_k) = \sum_{i=1}^k H(X_i|X_1,\ldots,X_{i-1}).$$

b) Chain rule for differential entropy follows from a similar argument, and the marginalization of the joint PDF, i.e.,

$$\int_{(x_{i+1},\ldots,x_k)\in\mathcal{X}^{k-i}} p_{X_1,\ldots,X_k}(x_1,\ldots,x_k)\ \mathsf{d}x_{i+1}\ldots\mathsf{d}x_k = p_{X_1,\ldots,X_i}(x_1,\ldots,x_i).$$

2) a) Using Radon-Nikodym chain rule, we get

$$D_{\mathsf{KL}}(P_{X,Y}\|Q_{X,Y}) = \mathbb{E}_{P_{X,Y}}\left[\log\left(\frac{\mathsf{d}P_{X,Y}}{\mathsf{d}Q_{X,Y}}\right)\right] = \mathbb{E}_{P_{X,Y}}\left[\log\left(\frac{\mathsf{d}P_X\mathsf{d}P_{Y|X}}{\mathsf{d}Q_X\mathsf{d}Q_{Y|X}}\right)\right].$$

Then,

$$\begin{aligned} D_{\mathsf{KL}}(P_{X,Y}\|Q_{X,Y}) &= \mathbb{E}_{P_{X,Y}}\left[\log\left(\frac{\mathsf{d}P_X}{\mathsf{d}Q_X}\right)\right] + \mathbb{E}_{P_{X,Y}}\left[\log\left(\frac{\mathsf{d}P_{Y|X}}{\mathsf{d}Q_{Y|X}}\right)\right] \\ &= \mathbb{E}_{P_X}\left[\log\left(\frac{\mathsf{d}P_X}{\mathsf{d}Q_X}\right)\right] + \mathbb{E}_{P_X}\left[D_{\mathsf{KL}}(P_{Y|X}(\cdot|X)\|Q_{Y|X}(\cdot|X))\right] \\ &= D_{\mathsf{KL}}(P_X\|Q_X) + D_{\mathsf{KL}}(P_{Y|X}\|Q_{Y|X}|P_X). \end{aligned}$$

b) Using the definition of mutual information and part (a), we have

$$I(X;Y) = D_{\mathsf{KL}}(P_{X,Y}\|P_X\otimes P_Y) = D_{\mathsf{KL}}(P_X\|P_X) + D_{\mathsf{KL}}(P_{Y|X}\|P_Y|P_X) = D_{\mathsf{KL}}(P_{Y|X}\|P_Y|P_X),$$

where the last equality follows from the fact that $D_{\mathsf{KL}}(P_X\|P_X) = \mathbb{E}_{P_X}[\log(1)] = 0$.

c) Using the fact that $\mathsf{d}P_{X,Y} = \mathsf{d}P_{Y,X}$, we get

$$I(Y;X) = D_{\mathsf{KL}}(P_{Y,X}\|P_Y \otimes P_X) = D_{\mathsf{KL}}(P_{X,Y}\|P_X \otimes P_Y) = I(X;Y).$$

d) Using part (b) and (a), we get

$$I(X;Y,Z) = D_{\mathsf{KL}}(P_{Y,Z|X}\|P_{Y,Z}|P_X) = D_{\mathsf{KL}}(P_{Y|X}\|P_Y|P_X) + D_{\mathsf{KL}}(P_{Z|X,Y}\|P_{Z|Y}|P_{X,Y}) \geq D_{\mathsf{KL}}(P_{Y|X}\|P_Y|P_X),$$

where the last step follows from the non-negativity of the KL-divergence. Thus, $I(X;Y,Z) \geq I(X;Y)$.

e) Consider the transition kernel induced by mapping (Id, f). We obtain $P_{X,f(Y)}$ and $P_{X,f(Y)}$ by passing $P_{X,Y}$ and $P_X \otimes P_Y$ through this kernel, respectively. Using the f-divergence DPI, we get

$$D_{\mathsf{KL}}(P_{X,Y}\|P_X \otimes P_Y) \geq D_{\mathsf{KL}}(P_{X,f(Y)}\|P_X \otimes P_{f(Y)}).$$

Thus, $I(X;Y) \geq I(X;f(Y))$. If f is one-to-one, then for discrete X, we have

$$I(X;f(X)) = H(f(X)) - H(f(X)|X) = H(f(X)) = H(X),$$

where the last equality follows from the fact that f is a bijection and because entropy is invariant to relabeling. For the continuous case, we show that $I(X;X) = \infty$. Assume $P_X \ll \lambda$ where λ is the Lebesgue measure. From the definition $I(X;X) = D_{\mathsf{KL}}(P_{XX}||P_X \otimes P_X)$. We will show that $P_{XX} \not\ll P_X \otimes P_X$, thereby implying that KL divergence diverges, as claimed. Define the diagonal set $\Delta := \{(x,x) : x \in \mathcal{X}\}$. Then,

$$\begin{aligned} P_{XX}(\Delta) &= \int_\Delta \mathsf{d}P_{XX}(x,x) = \int_\mathcal{X}\int_\mathcal{X} \mathbb{1}_{\{x=x'\}}\mathsf{d}P_{XX}(x,x) = \int_\mathcal{X} \mathsf{d}P_X(x)\int_\mathcal{X} \mathbb{1}_{\{x=x'\}}\mathsf{d}P_{X|X}(x'|x) \\ &= \int_\mathcal{X} \mathsf{d}P_X(x)\int_\mathcal{X} \mathbb{1}_{\{x=x'\}}\mathsf{d}\delta_x(x') = \int_\mathcal{X} \delta_x(x)\mathsf{d}P_X(x) = 1. \end{aligned}$$

However,

$$\begin{aligned} P_X \otimes P_X(\Delta) &= \int_\Delta \mathsf{d}P_X \otimes P_X(x,x') = \int_\mathcal{X}\int_\mathcal{X} \mathbb{1}_{\{x=x'\}}\mathsf{d}P_X \otimes P_X \\ &= \int_\mathcal{X} \mathsf{d}P_X(x)\int_\mathcal{X} \mathbb{1}_{\{x=x'\}}\mathsf{d}P_X(x') = \int_\mathcal{X} P_X(x)\mathsf{d}P_X(x) = 0, \end{aligned}$$

where the last equality follows from the fact that $P_X(x) = 0$ for all $x \in \mathcal{X}$ because $P_X \ll \lambda$. Thus, $P_{XX} \not\ll P_X \otimes P_X$ as $P_{XX}(\Delta) > 0$ while $P_X \otimes P_X(\Delta) = 0$.

3) a)

$$H(Z) \leq H(Z,X) = H(X) + H(X+Y|X) = H(X) + \sum_{x\in\mathcal{X}} P_X(x)H(x+Y|X=x).$$

Using the fact that entropy is invariant to relabeling, we have $H(x+Y|X=x) = H(Y|X=x)$ for all $x \in \mathcal{X}$. Then,

$$H(Z) \leq H(X) + \sum_{x\in\mathcal{X}} P_X(x)H(Y|X=x) = H(X) + H(Y|X) \leq H(X) + H(Y),$$

where the last equality follows from the fact that conditioning cannot increase entropy. The inequality

$\max\{H(X), H(Y)\} \leq H(Z)$ follows from part (b).

b) Using the fact that entropy is invariant to relabeling, we have

$$H(Z|X) = \sum_{x \in \mathcal{X}} P_X(x) H(x + Y|X = x) = \sum_{x \in \mathcal{X}} P_X(x) H(Y|X = x) = H(Y|X).$$

If X and Y are independent, we have $H(Y|X) = H(Y)$. Using the fact that conditioning cannot increase entropy, we get

$$H(Z) \geq H(Z|X) = H(Y|X) = H(Y).$$

Similarly, we get

$$H(Z) \geq H(Z|Y) = H(X|Y) = H(X).$$

c) Let X be an arbitrary random variable with $H(X) > 0$ and define $Y = -X$. Then, $Z = 0$ and $H(Z) = 0$.

4) a)

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z) \geq H(X|Z).$$

The equality holds if and only if $H(Y|X, Z) = 0$. That is, Y is a function of X and Z.

b)

$$I(X, Y; Z) = I(X; Z) + I(Y; Z|X) \geq I(X; Z).$$

The equality holds if and only if $I(Y; Z|X) = 0$, i.e., Y and Z are conditionally independent given X.

c) We have $H(X, Y, Z) - H(X, Y) = H(Z|X, Y)$. Also, $H(X, Z) - H(X) = H(Z|X)$. Using the fact that conditioning cannot increase entropy, we get

$$H(X, Y, Z) - H(X, Y) = H(Z|X, Y) \leq H(Z|X) = H(X, Z) - H(X).$$

The equality holds if $H(Z|X, Y) = H(Z|X)$, i.e., Y and Z are conditionally independent given X.

d) We have $I(X, Y; Z) = I(X; Z) + I(Y; Z|X) = I(Y; Z) + I(X; Z|X)$. Then,

$$I(X; Z|X) = I(X; Z) + I(Y; Z|X) - I(Y; Z).$$

The equality always holds.

5) a) Define $Q \in \mathcal{P}(\mathbb{N})$ with PMF $q(n) = 6/(\pi^2 n^2)$ for all $n \in \mathbb{N}$. Then,

$$H(P) = \mathbb{E}_P\left[\log\left(\frac{1}{p(X)}\right)\right] = \mathbb{E}_P\left[\log\left(\frac{q(X)}{p(X)}\right)\right] + \mathbb{E}_P\left[\log\left(\frac{1}{q(X)}\right)\right] = -D_{\mathsf{KL}}(P\|Q) + \mathbb{E}_P\left[\log\left(\frac{1}{q(X)}\right)\right].$$

Using the facct that $D_{\mathsf{KL}}(P\|Q) \geq 0$, we get

$$H(P) \leq \mathbb{E}_P\left[\log\left(\frac{\pi^2 X^2}{6}\right)\right] = \log\left(\frac{\pi^2}{6}\right) + 2\mathbb{E}_P\left[\log(X)\right].$$

b) Define $P \in \mathcal{P}(\mathbb{N})$ with PMF $p(n) = c/(n(\log n)^2)$ for all $n \in \{2, 3, \ldots\}$ where $c := \sum_{n=2}^{\infty} 1/(n(\log n)^2)$. We show that p is a valid PMF by proving that c is finite.

$$c \leq \int_2^\infty \frac{1}{x(\log x)^2} \mathsf{d}x = -\frac{1}{\log(x)}\Big|_2^\infty = \log 2.$$

Then,

$$H(P) = \sum_{n=2}^\infty \frac{c}{n(\log n)^2} \log(n(\log n)^2/c) = \sum_{n=2}^\infty \frac{c}{n \log n} - \log c + 2c \sum_{n=2}^\infty \frac{\log\log n}{n(\log n)^2} \geq \sum_{n=2}^\infty \frac{c}{n \log n} - \log 2.$$

$H(P) = \infty$ since $\sum_{n=2}^\infty 1/(n \log n) = \infty$.

6) a) The convexity follows from the fact that $(P, Q) \to D_{\mathsf{KL}}(P \| Q)$ is convex in (P, Q) and that $I(P_X, P_{Y|X}) = D_{\mathsf{KL}}\left(P_{Y|X} \| P_Y | P_X\right)$.

b) (i) By the total probability law, we get $P_X = \alpha P_X^{(2)} + (1-\alpha) P_X^{(1)}$.

(ii) We have $P_{X,Y,\Theta} = P_\Theta P_{X|\Theta} P_{Y|X,\Theta} = P_\Theta P_{X|\Theta} P_{Y|X}$ where the equality follows from the fact that Y is obtained by passing X through the kernel $P_{Y|X}$.

(iii) Using part (b), we have

$$I(X;Y) = I(X,\Theta;Y) = I(\Theta;Y) + I(X;Y|\Theta) \geq I(X;Y|\Theta).$$

Thus, for all $\alpha \in [0,1]$

$$I(\alpha P_X^{(2)} + (1-\alpha) P_X^{(1)}, P_{Y|X}) \geq \alpha I(P_X^{(2)}, P_{Y|X}) + (1-\alpha) I(P_X^{(1)}, P_{Y|X}).$$

7) Using chain rule, we get

$$I(X_1; X_2, \ldots, X_n) = I(X_1; X_2) + \sum_{k=3}^n I(X_k; X_1 | X_2, \ldots, X_{k-1})$$

Using the fact that $\{Z_i\}_{i=1}^\infty$ is a sequence of i.i.d. random variables, we get $X_1 \to X_2 \ldots \to X_n$. Thus, for all $k \geq 3$, we have

$$I(X_k; X_1 | X_2, \ldots, X_{k-1}) = 0.$$

Then,

$$I(X_1; X_2, \ldots, X_n) = I(X_1; X_2) = I(Z_1 + Z_2; Z_1) = H(X_2) - H(Z_2).$$

We have

$$p_{X_2}(x) = \begin{cases} 1/2, & x = 1, \\ 1/4, & x = 0, 2. \end{cases}$$

Then, $H(X_2) = 3/2$. Also, $H(Z_2) = 1$ as $Z_2 \sim \mathrm{Ber}(1/2)$. Thus,

$$I(X_1; X_2, \ldots, X_n) = I(X_1; X_2) = \frac{1}{2}.$$

8) Using the bounds on the PDFs, for all $x \in [0,1]$. we have

$$0 < \frac{c_1}{c_2} \leq \frac{q(x)}{p(x)} \leq \frac{c_2}{c_1} < \infty.$$

Note that $c_1/c_2 < 1 < c_2/c_1$. Using Taylor's theorem, for all $y \in [c_1/c_2, c_2/c_1]$ there exists a $\varepsilon(y)$ such that $|\varepsilon(y) - 1| \leq |y-1|$ for which it holds that

$$\log(y) = y - 1 - \frac{1}{2\varepsilon(y)^2}(y-1)^2. \tag{1}$$

Then, for all $y \in [c_1/c_2, c_2/c_1]$, we have

$$y - 1 - \frac{c_2^2}{2c_1^2}(y-1)^2 \leq \log(y) \leq y - 1 - \frac{c_1^2}{2c_2^2}(y-1)^2. \tag{2}$$

Let $y = q(x)/p(x)$. Then, using the definition of KL-divergence, we get

$$\begin{aligned} D_{\mathsf{KL}}(P\|Q) &= -\int_0^1 p(x)\frac{q(x)}{p(x)}\mathsf{d}x \\ &\leq -\int_0^1 p(x)\left(\frac{q(x)}{p(x)} - 1 - \frac{c_2^2}{2c_1^2}\frac{(q(x)-p(x))^2}{p(x)^2}\right)\mathsf{d}x \\ &= -\int_0^1 p(x)\mathsf{d}x + 1 + \frac{c_2^2}{2c_1^2}\int_0^1 \frac{(q(x)-p(x))^2}{p(x)}\mathsf{d}x \\ &\leq \frac{c_2^2}{2c_1^3}\int_0^1 (q(x)-p(x))^2\mathsf{d}x. \end{aligned}$$

Setting $k_2 = c_2^2/(2c_1^3)$. The lower bound can be derived similarly using the upper bound in Inequality (2) with $k_1 = c_1^2/(2c_2^3)$.