

Problem 1

Let \mathcal{X} be a finite set and p_X be a probability distribution or a probability mass function (PMF) on \mathcal{X} . The Shannon entropy of p_X is defined as

$$H(p_X) \triangleq - \sum_{x \in \mathcal{X}} p_X(x) \log p_X(x)$$

1. Prove $\log x \leq x - 1$ and $\log \frac{1}{x} \geq 1 - x$ for all $x > 0$.
2. $\sum_{x \in \mathcal{X}} p_X(x) \log \frac{1}{p_X(x)} \leq \log |\mathcal{X}|$
3. $H(X) + H(Y) \geq H(X, Y)$ where $H(X, Y) = H(p_{X,Y})$ is the entropy of a joint PMF, $H(X) = H(p_X)$ where p_X is marginal of $p_{X,Y}$

Solution:

1. We have $\log x = \int_1^x \frac{1}{t} dt$ and $x - 1 = \int_1^x dt$. Now for $x \geq 1$ for all $t \geq 1$ we have $1 \geq \frac{1}{t}$. Hence

$$\int_1^x \frac{1}{t} dt \leq \int_1^x dt \iff \log x \leq x - 1$$

For $0 < x < 1$ we have $t < 1$ hence $\frac{1}{t} \geq 1$. Hence

$$\int_x^1 \frac{1}{t} dt \geq \int_x^1 dt \iff -\log x \geq 1 - x \iff x - 1 \geq \log x$$

Therefore $\forall x > 0$ we have $\log x \leq x - 1$.

Now we have $\log x \leq x - 1 \iff 1 - x \leq -\log x \iff 1 - x \leq \log \frac{1}{x}$.

2.

$$\begin{aligned} \sum_{x \in \mathcal{X}} p_X(x) \log \frac{1}{p_X(x)} - \log |\mathcal{X}| &= \sum_{x \in \mathcal{X}} p_X(x) \log \frac{1}{p_X(x)} - \sum_{x \in \mathcal{X}} p_X(x) \log |\mathcal{X}| \\ &= \sum_{x \in \mathcal{X}} p_X(x) \log \frac{1}{|\mathcal{X}| p_X(x)} \\ &\leq \sum_{x \in \mathcal{X}} p_X(x) \left[\frac{1}{|\mathcal{X}| p_X(x)} - 1 \right] \quad [\text{Using Part (1)}] \\ &= \sum_{x \in \mathcal{X}} \left[\frac{1}{|\mathcal{X}|} - p_X(x) \right] = 1 - 1 = 0 \end{aligned}$$

Hence we get

$$\sum_{x \in \mathcal{X}} p_X(x) \log \frac{1}{p_X(x)} - \log |\mathcal{X}| \iff \sum_{x \in \mathcal{X}} p_X(x) \log \frac{1}{p_X(x)} \leq \log |\mathcal{X}|$$

Now if we take the base of the log any number the inequality still holds since the multiplicative factor due to change of basis gets canceled out.

3.

$$\begin{aligned}
H(X) + H(Y) - H(X, Y) &= - \sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) - \sum_{y \in \mathcal{Y}} p_Y(y) \log p_Y(y) \\
&\quad + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) \log p_{XY}(x, y) \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_X(x) - \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p_{XY}(x, y) \log p_Y(y) \\
&\quad + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) \log p_{XY}(x, y) \\
&= - \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) \log \frac{p_X(x)p_Y(y)}{p_{XY}(x, y)} \\
&= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \\
&\geq \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) \left[1 - \frac{p_X(x)p_Y(y)}{p_{XY}(x, y)} \right] \quad \text{[Using Part (1)]} \\
&= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) - \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) \frac{p_X(x)p_Y(y)}{p_{XY}(x, y)} \\
&= 1 - \sum_{x \in \mathcal{X}} p_X(x) \left[\sum_{y \in \mathcal{Y}} p_Y(y) \right] \\
&= 1 - \sum_{x \in \mathcal{X}} p_X(x) = 1 - 1 = 0
\end{aligned}$$

Hence we got $H(X) + H(Y) \geq H(X, Y)$.

□

Problem 2

Let $p_X(x)$ be a PMF on \mathcal{X} . For $n \in \mathbb{N}$, $\delta > 0$, let

$$T_\delta^n(p_X) \triangleq \left\{ x^n \in \mathcal{X}^n \mid \left| \frac{N(a|x^n)}{n} - p_X(a) \right| \leq \frac{\delta p_X(a)}{\log |\mathcal{X}|} \forall a \in \mathcal{X} \right\}$$

where $N(a|x^n) = \sum_{i=1}^n \mathbb{1}_{\{x_i=a\}}$ denotes the number of occurrences of a in the sequences $x_1 x_2 \cdots x_n$.

1. Prove that

$$\sum_{x^n \notin T_\delta^n(p_X)} \prod_{i=1}^n p_X(x_i) \leq \exp \left[-\frac{2n\delta^2 \eta_{p_X}^2}{(\log |\mathcal{X}|)^2} \right]$$

where $\eta_{p_X} = \min_{a \in \mathcal{X}} \{p_X(a) \mid 0 < p_X(a) < 1\}$

2. Prove that

$$\left[1 - \exp \left(-\frac{2n\delta^2 \eta_{p_X}^2}{(\log |\mathcal{X}|)^2} \right) \right] \exp[n(H(p_X) - \delta)] \leq |T_\delta^n(p_X)| \leq \exp[n(H(p_X) + \delta)]$$

3. Prove that

$$x^n \in T_\delta^n(p_X) \implies \exp[-n(H(p_X) + \delta)] \leq \prod_{i=1}^n p_X(x_i) \leq \exp[-n(H(p_X) - \delta)]$$

Solution:

1. $\sum_{x^n \notin T_\delta^n(p_X)} \prod_{i=1}^n p_X(x_i) = \sum_{x^n \notin T_\delta^n(p_X)} p_X^n(x^n) = \Pr[x^n \notin T_\delta^n(p_X)]$. If $x^n \notin T_\delta^n(p_X)$ then there exists $a \in \mathcal{X}$ such that $\left| \frac{N(a|x^n)}{n} - p_X(a) \right| > \frac{\delta p_X(a)}{\log |\mathcal{X}|}$. Now $N(a|x^n) = \sum_{i=1}^n \mathbb{1}_{x_i=a}$. Hence take the indicator random variables $\mathbb{1}_{x_i=a}$ for $a, i \in [n]$ then $\mathbb{E}[\mathbb{1}_{x_i=a}] = p_X(a)$. Then by Hoeffding Inequality we get

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_i=a} - p_X(a) \right| > \frac{\delta p_X(a)}{\log |\mathcal{X}|} \right] \leq 2 \exp \left[-2n \left(\frac{\delta p_X(a)}{\log |\mathcal{X}|} \right)^2 \right] \leq 2 \exp \left[-2n \left(\frac{\delta \eta_{p_X}}{\log |\mathcal{X}|} \right)^2 \right]$$

So

$$\begin{aligned} \Pr[x^n \notin T_\delta^n(p_X)] &\leq \sum_{a \in \mathcal{X}} \Pr \left[\left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_i=a} - p_X(a) \right| > \frac{\delta p_X(a)}{\log |\mathcal{X}|} \right] \\ &\leq \sum_{a \in \mathcal{X}} 2 \exp \left[-2n \left(\frac{\delta p_X(a)}{\log |\mathcal{X}|} \right)^2 \right] \leq 2 \exp \left[-2n \left(\frac{\delta \eta_{p_X}}{\log |\mathcal{X}|} \right)^2 \right] \\ &= 2|\mathcal{X}| \exp \left[-\frac{2n\delta^2 \eta_{p_X}^2}{\log^2 |\mathcal{X}|} \right] \end{aligned}$$

2. Using part (3) of we have

$$1 \geq \sum_{x^n \in T_\delta^n(p_X)} p_X^n(x^n) \geq \sum_{x^n \in T_\delta^n(p_X)} \exp[-n(H(p_X) + \delta)] \geq |T_\delta^n(p_X)| \exp[-n(H(p_X) + \delta)]$$

Therefore we obtain

$$|T_\delta^n(p_X)| \leq \exp[n(H(p_X) + \delta)]$$

Now

$$Pr[x^n \notin T_\delta^n(p_X)] \leq 2|\mathcal{X}| \exp \left[-\frac{2n\delta^2\eta_{p_X}^2}{\log^2 |\mathcal{X}|} \right] \implies Pr[x^n \in T_\delta] \geq 1 - 2|\mathcal{X}| \exp \left[-\frac{2n\delta^2\eta_{p_X}^2}{\log^2 |\mathcal{X}|} \right]$$

And again using part (3)

$$Pr[x^n \in T_\delta] = \sum_{x^n \in T_\delta^n(p_X)} p_X^n(x^n) \leq \sum_{x^n \in T_\delta^n(p_X)} \exp[-n(H(p_X) - \delta)] \leq |T_\delta^n(p_X)| \exp[-n(H(p_X) - \delta)]$$

Therefore we have

$$|T_\delta^n(p_X)| \geq \left[1 - 2|\mathcal{X}| \exp \left(-\frac{2n\delta^2\eta_{p_X}^2}{\log^2 |\mathcal{X}|} \right) \right] \exp[n(H(p_X) - \delta)]$$

Hence we finally obtain

$$\left[1 - 2|\mathcal{X}| \exp \left(\frac{2n\delta^2\eta_{p_X}^2}{(\log |\mathcal{X}|)^2} \right) \right] \exp[n(H(p_X) - \delta)] \leq |T_\delta^n(p_X)| \leq \exp[n(H(p_X) + \delta)]$$

3. $p_X(x^n) = \prod_{i=1}^n p_X(x_i) = \prod_{a \in \mathcal{X}} p_X(a)^{N(a|x^n)}$. Now from the definition we get for all $a \in \mathcal{X}$ if $x^n \in T_\delta^n(p_X)$

$$-\frac{\delta p_X(a)}{\log |\mathcal{X}|} \leq \frac{N(a|x^n)}{n} - p_X(a) \leq \frac{\delta p_X(a)}{\log |\mathcal{X}|} \implies np_X(a) \left[1 - \frac{\delta}{\log |\mathcal{X}|} \right] \leq N(a|x^n) \leq np_X(a) \left[1 + \frac{\delta}{\log |\mathcal{X}|} \right]$$

Now we get

$$\begin{aligned} \prod_{a \in \mathcal{X}} p_X(a)^{N(a|x^n)} &\leq \prod_{a \in \mathcal{X}} p_X(a)^{np_X(a) \left[1 - \frac{\delta}{\log |\mathcal{X}|} \right]} \\ &= \prod_{x \in \mathcal{X}} \exp[np_X(a) \left[1 - \frac{\delta}{\log |\mathcal{X}|} \right] \log p_X(a)] \\ &= \exp \left[\sum_{x \in \mathcal{X}} np_X(a) \left(1 - \frac{\delta}{\log |\mathcal{X}|} \right) \log p_X(a) \right] \\ &= \exp \left[n \left(1 - \frac{\delta}{\log |\mathcal{X}|} \right) \sum_{x \in \mathcal{X}} p_X(a) \log p_X(a) \right] \\ &= \exp \left[-n \left(1 - \frac{\delta}{\log |\mathcal{X}|} \right) H(p_X) \right] \end{aligned}$$

Similarly we get

$$\prod_{a \in \mathcal{X}} p_X(a)^{N(a|x^n)} \geq \exp \left[-n \left(1 + \frac{\delta}{\log |\mathcal{X}|} \right) H(p_X) \right]$$

By Problem 1.(2) we have $H(p_X) \leq \log |\mathcal{X}|$. Hence

$$\begin{aligned} -n \left(H(p_X) + \frac{\delta H(p_X)}{\log |\mathcal{X}|} \right) &\geq -n(H(p_X) + \delta) \\ -n \left(H(p_X) - \frac{\delta H(p_X)}{\log |\mathcal{X}|} \right) &\leq -n(H(p_X) - \delta) \end{aligned}$$

Therefore we get

$$\exp[-n(H(p_X) + \delta)] \leq \prod_{i=1}^n p_X(x_i) \leq \exp[-n(H(p_X) - \delta)]$$

□

Definitions: Let $p_{X,Y}$ be a joint PMF on $\mathcal{X} \times \mathcal{Y}$ where \mathcal{X}, \mathcal{Y} are finite sets. (Essentially $p_{X,Y}(x,y) \geq 0$ and $\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y) = 1$). We define the marginal of $p_{X,Y}$ on X as $p_X(x) \triangleq \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y)$ for $x \in \mathcal{X}$ and marginal of $p_{X,Y}$ on Y as $p_Y(y) \triangleq \sum_{x \in \mathcal{X}} p_{X,Y}(x,y)$ for $y \in \mathcal{Y}$.

For a pair $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ of sequences we define $N(a,b | x^n, y^n) = \sum_{i=1}^n \mathbb{1}_{\{(x_i, y_i) = (a,b)\}}$ as the number of occurrences of (a,b) in (x^n, y^n) .

Next the joint typical set wrt $p_{X,Y}$ is defined as

$$T_\delta^n(p_{X,Y}) \triangleq \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \mid \left| \frac{N(a,b | x^n, y^n)}{n} - p_{X,Y}(a,b) \right| \leq \frac{\delta p_{X,Y}(a,b)}{\log |\mathcal{X}| |\mathcal{Y}|} \forall (a,b) \in \mathcal{X} \times \mathcal{Y} \right\}$$

Problem 3

1. Prove that if $p_{X,Y}(a,b) = 0$ for some $(a,b) \in \mathcal{X} \times \mathcal{Y}$ and $(x^n, y^n) \in T_\delta^n(p_{X,Y})$ then $N(a,b | x^n, y^n) = 0$. In other words, a pair that has 0 probability does not occur in any typical pair of sequences.
2. Let $\eta_{p_{X,Y}} = \min_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \{p_{X,Y}(x,y) \mid 0 < p_{X,Y}(x,y) < 1\}$. Use the Hoeffding Inequality to prove that

$$\sum_{(x^n, y^n) \notin T_\delta^n(p_{X,Y})} p_{X,Y}^n(x^n, y^n) \leq 2|\mathcal{X}| |\mathcal{Y}| \exp \left[-\frac{2n\delta^2 \eta_{p_{X,Y}}^2}{(\log |\mathcal{X}| |\mathcal{Y}|)^2} \right]$$

Hoeffding Inequality: Let Z_1, \dots, Z_m are independent and identically distributed random variables for which $P[a \leq Z_i \leq b] = 1$ for ever $1 \leq i \leq m$ and $\mu = \mathbb{E}[Z_i]$. Then for every $\epsilon > 0$

$$\Pr \left[\left| \frac{1}{m} \sum_{i=1}^m Z_i - \mu \right| > \epsilon \right] \leq 2 \exp \left[-2m \frac{\epsilon^2}{(b-a)^2} \right]$$

3. For any $(x^n, y^n) \in T_\delta^n(p_{X,Y})$ prove that

$$2^{-n[H(p_{X,Y}) + \delta]} \leq p_{X,Y}^n(x^n, y^n) = \prod_{i=1}^n p_{X,Y}(x_i, y_i) \leq 2^{-n[H(p_{X,Y}) - \delta]}$$

4. Prove that

$$(1 - \tilde{\delta}) 2^{n[H(p_{X,Y}) - \delta]} \leq |T_\delta^n(p_{X,Y})| \leq 2^{n[H(p_{X,Y}) + \delta]}$$

$$\text{where } \tilde{\delta} = 2|\mathcal{X}| |\mathcal{Y}| \exp \left[-\frac{2n\delta^2 \eta_{p_{X,Y}}^2}{(\log |\mathcal{X}| |\mathcal{Y}|)^2} \right]$$

5. Prove that $(x^n, y^n) \in T_\delta^n(p_{X,Y})$ then $x^n \in T_\delta^n(p_X)$ and $y^n \in T_\delta^n(p_Y)$.

Solution:

1. Given that $p_{X,Y}(a,b) = 0$. Now if $(x^n, y^n) \in T_\delta^n(p_{X,Y})$

$$\left| \frac{N(a,b | x^n, y^n)}{n} - p_{X,Y}(a,b) \right| \leq \frac{\delta p_{X,Y}(a,b)}{\log |\mathcal{X}| |\mathcal{Y}|}$$

putting the given value $p_{X,Y}(a,b) = 0$ we get

$$\left| \frac{N(a,b | x^n, y^n)}{n} \right| \leq 0$$

Hence we get $\frac{N(a,b | x^n, y^n)}{n} = 0 \iff N(a,b | x^n, y^n) = 0$.

2. $\sum_{(x^n, y^n) \notin T_\delta^n(p_{XY})} p_{XY}^n(x^n, y^n) = \Pr[(x^n, y^n) \notin T_\delta^n(p_{XY})]$. If $(x^n, y^n) \notin T_\delta^n(p_{XY})$ then there exists $(a, b) \in \mathcal{X} \times \mathcal{Y}$ such that

$$\left| \frac{N(a, b | x^n, y^n)}{n} - p_{XY}(a, b) \right| > \frac{\delta p_{XY}(a, b)}{\log |\mathcal{X}| |\mathcal{Y}|}$$

Now we have $N(a, b | x^n, y^n) = \sum_{i=1}^n \mathbb{1}_{\{(x_i, y_i) = (a, b)\}}$. Take the indicator random variables $\mathbb{1}_{(x_i, y_i) = (a, b)}$ for $(a, b) \in \mathcal{X} \times \mathcal{Y}$ for each $i \in [n]$. Then $\mathbb{E} [\mathbb{1}_{(x_i, y_i) = (a, b)}] = p_{XY}(a, b)$. Hence by Hoeffding Inequality

$$\begin{aligned} \Pr \left[\left| \frac{1}{n} \sum_{(a, b) \in \mathcal{X} \times \mathcal{Y}} \mathbb{1}_{(x_i, y_i) = (a, b)} - p_{XY}(a, b) \right| > \frac{\delta p_{XY}(a, b)}{\log |\mathcal{X}| |\mathcal{Y}|} \right] &\leq 2 \exp \left[-2n \left(\frac{\delta p_{XY}(a, b)}{\log |\mathcal{X}| |\mathcal{Y}|} \right)^2 \right] \\ &\leq 2 \exp \left[-\frac{2n \delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right] \end{aligned}$$

So by union bound we get

$$\begin{aligned} \Pr[(x^n, y^n) \notin T_\delta^n(p_{XY})] &\leq \sum_{(a, b) \in \mathcal{X} \times \mathcal{Y}} \Pr \left[\left| \frac{1}{n} \sum_{(a, b) \in \mathcal{X} \times \mathcal{Y}} \mathbb{1}_{(x_i, y_i) = (a, b)} - p_{XY}(a, b) \right| > \frac{\delta p_{XY}(a, b)}{\log |\mathcal{X}| |\mathcal{Y}|} \right] \\ &\leq \sum_{(a, b) \in \mathcal{X} \times \mathcal{Y}} 2 \exp \left[-\frac{2n \delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right] = 2 |\mathcal{X}| |\mathcal{Y}| \exp \left[-\frac{2n \delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right] \end{aligned}$$

Therefore we get

$$\sum_{(x^n, y^n) \notin T_\delta^n(p_{XY})} p_{XY}^n(x^n, y^n) \leq 2 |\mathcal{X}| |\mathcal{Y}| \exp \left[-\frac{2n \delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right]$$

3. $p_{XY}^n(x^n, y^n) = \prod_{i=1}^n p_{XY}(x_i, y_i) = \prod_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(a, b)^{N(a, b | x^n, y^n)}$. Now from the definition of $T_\delta^n(p_{XY})$ we get

$$n p_{XY}(a, b) \left[1 - \frac{\delta}{\log |\mathcal{X}| |\mathcal{Y}|} \right] \leq N(a, b | x^n, y^n) \leq n p_{XY}(a, b) \left[1 + \frac{\delta}{\log |\mathcal{X}| |\mathcal{Y}|} \right]$$

So we have

$$\begin{aligned} \prod_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(a, b)^{N(a, b | x^n, y^n)} &\leq \prod_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(a, b)^{n p_{XY}(a, b) \left[1 - \frac{\delta}{\log |\mathcal{X}| |\mathcal{Y}|} \right]} \\ &= \prod_{(a, b) \in \mathcal{X} \times \mathcal{Y}} 2^{n p_{XY}(a, b) \left(1 - \frac{\delta}{\log |\mathcal{X}| |\mathcal{Y}|} \right) \log p_{XY}(a, b)} \\ &= 2^{\sum_{(a, b) \in \mathcal{X} \times \mathcal{Y}} n p_{XY}(a, b) \left(1 - \frac{\delta}{\log |\mathcal{X}| |\mathcal{Y}|} \right) \log p_{XY}(a, b)} \\ &= 2^{n \left(1 - \frac{\delta}{\log |\mathcal{X}| |\mathcal{Y}|} \right) \sum_{(a, b) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(a, b) \log p_{XY}(a, b)} \\ &= 2^{-n \left(1 - \frac{\delta}{\log |\mathcal{X}| |\mathcal{Y}|} \right) H(p_{XY})} \end{aligned}$$

Similarly we obtain

$$\prod_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p_{XY}(a, b)^{N(a, b | x^n, y^n)} \geq 2^{-n \left(1 + \frac{\delta}{\log |\mathcal{X}| |\mathcal{Y}|} \right) H(p_{XY})}$$

Now we will prove a claim

Claim: $H(p_{XY}) \leq \log |\mathcal{X}| |\mathcal{Y}|$

Proof:

$$\begin{aligned}
& \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log \frac{1}{p_{XY}(x, y)} - \log |\mathcal{X}| |\mathcal{Y}| \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log \frac{1}{p_X(x)} - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log |\mathcal{X}| \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log \frac{1}{(|\mathcal{X}| |\mathcal{Y}|) p_{XY}(x, y)} \\
&\leq \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \left[\frac{1}{(|\mathcal{X}| |\mathcal{Y}|) p_{XY}(x, y)} - 1 \right] \quad [\text{Using 1.(1)}] \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left[\frac{1}{|\mathcal{X}| |\mathcal{Y}|} - p_{XY}(x, y) \right] = 1 - 1 = 0
\end{aligned}$$

□

Now using the claim we get

$$\begin{aligned}
2^{-n(H(p_{XY}) - \frac{\delta H(p_{XY})}{\log |\mathcal{X}| |\mathcal{Y}|})} &\leq 2^{-n(H(p_{XY}) - \delta)} \\
2^{-n(H(p_{XY}) + \frac{\delta H(p_{XY})}{\log |\mathcal{X}| |\mathcal{Y}|})} &\geq 2^{-n(H(p_{XY}) + \delta)}
\end{aligned}$$

Hence we get if $(x^n, y^n) \in T_\delta^n(p_{XY})$ then

$$2^{-n(H(p_{XY}) + \delta)} \leq p_{XY}^n(x^n, y^n) \leq 2^{-n(H(p_{XY}) - \delta)}$$

4. Using part (2) we have

$$1 \geq \sum_{(x^n, y^n) \in T_\delta^n(p_{XY})} p_{XY}^n(x^n, y^n) \geq \sum_{(x^n, y^n) \in T_\delta^n(p_{XY})} 2^{-n(H(p_{XY}) + \delta)} \geq |T_\delta^n(p_{XY})| 2^{-n(H(p_{XY}) + \delta)}$$

Hence we get

$$|T_\delta^n(p_{XY})| \leq 2^{n(H(p_{XY}) + \delta)}$$

In part (1) we proved $Pr[(x^n, y^n) \notin T_\delta^n(p_{XY})] \leq 2|\mathcal{X}| |\mathcal{Y}| \exp \left[-\frac{2n\delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right]$. Hence

$$Pr[(x^n, y^n) \in T_\delta(p_{XY})] \geq 1 - 2|\mathcal{X}| |\mathcal{Y}| \exp \left[-\frac{2n\delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right]$$

and

$$\begin{aligned}
Pr[(x^n, y^n) \in T_\delta^n(p_{XY})] &= \sum_{(x^n, y^n) \in T_\delta^n(p_{XY})} p_{XY}^n(x^n, y^n) \\
&\leq \sum_{(x^n, y^n) \in T_\delta^n(p_{XY})} 2^{-n(H(p_{XY}) - \delta)} \leq |T_\delta^n(p_{XY})| 2^{-n(H(p_{XY}) - \delta)}
\end{aligned}$$

Therefore we get

$$\begin{aligned}
|T_\delta^n(p_{XY})| 2^{-n(H(p_{XY}) - \delta)} &\geq 1 - 2|\mathcal{X}| |\mathcal{Y}| \exp \left[-\frac{2n\delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right] \\
\Rightarrow |T_\delta^n(p_{XY})| &\geq \left[1 - 2|\mathcal{X}| |\mathcal{Y}| \exp \left(-\frac{2n\delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right) \right] 2^{n(H(p_{XY}) - \delta)}
\end{aligned}$$

Therefore finally we get

$$\left[1 - 2|\mathcal{X}| |\mathcal{Y}| \exp \left(-\frac{2n\delta^2 \eta_{XY}^2}{\log^2 |\mathcal{X}| |\mathcal{Y}|} \right) \right] 2^{n(H(p_{XY}) - \delta)} \leq |T_\delta^n(p_{XY})| \leq 2^{n(H(p_{XY}) + \delta)}$$

5. Since $(x^n, y^n) \in T_\delta^n(p_{XY})$, for all $(a, b) \in \mathcal{X} \times \mathcal{Y}$

$$\left| \frac{N(a, b | x^n, y^n)}{n} - p_{XY}(a, b) \right| \leq \frac{\delta p_{XY}(a, b)}{\log |\mathcal{X}| |\mathcal{Y}|}$$

We have

$$N(a|x^n) = \sum_{b \in \mathcal{Y}} N(a, b|x^n, y^n) \text{ and } N(b|y^n) = \sum_{a \in \mathcal{X}} N(a, b|x^n, y^n)$$

Now

$$\begin{aligned} \left| \frac{N(a|x^n)}{n} - p_X(a) \right| &= \left| \frac{1}{n} \sum_{b \in \mathcal{Y}} N(a, b|x^n, y^n) - \sum_{b \in \mathcal{Y}} p_{XY}(a, b) \right| \\ &= \left| \sum_{b \in \mathcal{Y}} \left[\frac{N(a, b|x^n, y^n)}{n} - p_{XY}(a, b) \right] \right| \\ &\leq \sum_{b \in \mathcal{Y}} \left| \frac{N(a, b|x^n, y^n)}{n} - p_{XY}(a, b) \right| \\ &\leq \sum_{b \in \mathcal{Y}} \frac{\delta p_{XY}(a, b)}{\log |\mathcal{X}|} = \frac{\delta}{\log |\mathcal{X}|} \sum_{b \in \mathcal{Y}} p_{XY}(a, b) = \frac{\delta p_X(a)}{\log |\mathcal{X}|} \end{aligned}$$

Hence $x^n \in T_\delta^n(p_X)$.

$$\begin{aligned} \left| \frac{N(b|y^n)}{n} - p_Y(b) \right| &= \left| \frac{1}{n} \sum_{a \in \mathcal{X}} N(a, b|x^n, y^n) - \sum_{a \in \mathcal{X}} p_{XY}(a, b) \right| \\ &= \left| \sum_{a \in \mathcal{X}} \left[\frac{N(a, b|x^n, y^n)}{n} - p_{XY}(a, b) \right] \right| \\ &\leq \sum_{a \in \mathcal{X}} \left| \frac{N(a, b|x^n, y^n)}{n} - p_{XY}(a, b) \right| \\ &\leq \sum_{a \in \mathcal{X}} \frac{\delta p_{XY}(a, b)}{\log |\mathcal{X}|} = \frac{\delta}{\log |\mathcal{X}|} \sum_{a \in \mathcal{X}} p_{XY}(a, b) = \frac{\delta p_Y(b)}{\log |\mathcal{X}|} \end{aligned}$$

Hence $y^n \in T_\delta^n(p_Y)$.

□

Definitions: Suppose p_{XY} is a probability distribution (probability mass function (PMF)) on $\mathcal{X} \times \mathcal{Y}$. We recall the condition distribution $p_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)}$ and for a pair $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ of sequence $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ of sequences $p_{Y|X}^n(y^n|x^n) = \prod_{i=1}^n p_{Y|X}(y_i|x_i)$

We define

$$H(Y|X = x) \triangleq H(p_{Y|X}|X = x) = - \sum_{y \in \mathcal{Y}} p_{Y|X}(y|x) \log p_{Y|X}(y|x)$$

and

$$H(Y|X) = H(p_{Y|X}|p_X) \triangleq \sum_{x \in \mathcal{X}} p_X(x) h(Y|X = x)$$

For any $x^n \in \mathcal{X}^n$ define the conditional typical set of x^n as

$$T_\delta^n(p_{Y|X}|x^n) = \{y^n \in \mathcal{Y}^n \mid (x^n, y^n) \in T_\delta^n(p_{XY})\}$$

Problem 4

1. Prove that $\sum_{y \in \mathcal{Y}} p_{Y|X}(y|x) = 1$
2. Prove that $H(Y|X) = H(X, Y) - H(X)$
3. Prove that $H(Y|X) \geq 0$
4. Prove that Verify that if $x^n \notin T_\delta^n(p_X)$ then $T_\delta^n(p_{XY}|x^n) = \emptyset$
5. Suppose $x^n \in T_\delta^n(p_X)$ and $y^n \in T_\delta^n(p_{XY}|x^n)$ prove that

$$2^{-n[H(Y|X)+2\delta]} \leq p_{Y|X}^n(y^n|x^n) \leq 2^{-n[H(Y|X)-2\delta]}$$

6. Prove that if $x^n \in T_\delta^n(p_X)$ then

$$\sum_{y^n \in T_{2\delta}^n(p_{XY}|x^n)} p_{Y|X}^n(y^n|x^n) \geq 1 - 2|\mathcal{X}||\mathcal{Y}| \exp \left[-\frac{2n\delta^2}{(\log |\mathcal{X}||\mathcal{Y}|)^2} \eta_{p_{Y|X}} \right]$$

where $\eta_{p_{Y|X}} = \min_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \{p_{Y|X}(y|x) \mid 0 < p_{Y|X}(y|x) < 1\}$

7. Suppose $x^n \in T_\delta^n(p_X)$ then

$$(1 - \tilde{\delta})2^{n[H(Y|X)-4\delta]} \leq |T_\delta^n(p_{XY}|x^n)| \leq 2^{n[H(Y|X)+4\delta]}$$

where $\tilde{\delta} = 2|\mathcal{X}||\mathcal{Y}| \exp \left[-\frac{2n\delta^2}{(\log |\mathcal{X}||\mathcal{Y}|)^2} \eta_{p_{Y|X}} \right]$

Solution:

1. We have $p_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)}$. Hence

$$\begin{aligned} \sum_{y \in \mathcal{Y}} p_{Y|X}(y|x) &= \sum_{y \in \mathcal{Y}} \frac{p_{XY}(x,y)}{p_X(x)} \\ &= \frac{\sum_{y \in \mathcal{Y}} p_{XY}(x,y)}{p_X(x)} = \frac{p_X(x)}{p_X(x)} = 1 \end{aligned}$$

Therefore

$$\sum_{y \in \mathcal{Y}} p_{Y|X}(y|x) = 1$$

2.

$$\begin{aligned}
H(Y|X) &= - \sum_{x \in \mathcal{X}} p_X(x) H(Y|X=x) = - \sum_{x \in \mathcal{X}} p_X(x) \sum_{y \in \mathcal{Y}} p_{Y|X}(y|x) \log p_{Y|X}(y|x) \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_X(x) p_{Y|X}(y|x) \log p_{Y|X}(y|x) \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_{Y|X}(y|x)
\end{aligned}$$

Now we will show that $H(X, Y) = H(Y|X) + H(X)$.

$$\begin{aligned}
H(X, Y) &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_{XY}(x, y) \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log [p_{Y|X}(y|x) p_X(x)] \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) [\log p_{Y|X}(y|x) + \log p_X(x)] \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_{Y|X}(y|x) - \sum_{x \in \mathcal{X}} \left[\sum_{y \in \mathcal{Y}} p_{XY}(x, y) \right] \log p_X(x) \\
&= H(Y|X) - \sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) = H(Y|X) + H(X)
\end{aligned}$$

Hence we get $H(Y|X) = H(X, Y) - H(X)$

3. $p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)} = \frac{p_{XY}(x, y)}{\sum_{y \in \mathcal{Y}} p_{XY}(x, y)} \leq 1$ and $p_{Y|X}(y|x) \geq 0$. Now in the previous part we showed that

$$H(Y|X) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_{Y|X}(y|x)$$

Now

$$p_{Y|X}(y|x) \leq 1 \implies \log p_{Y|X}(y|x) \leq 0 \implies -\log p_{Y|X}(y|x) \geq 0$$

Hence

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y} \quad p_{XY}(x, y) \log p_{Y|X}(y|x) \geq 0$$

Therefore

$$H(Y|X) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_{Y|X}(y|x) \geq 0$$

4. By Problem 3.(5) we have if $(x^n, y^n) \in T_\delta^n(p_{XY})$ then $x^n \in T_\delta^n(p_X)$ and $y^n \in T_\delta^n(p_Y)$. Hence if $x^n \notin T_\delta^n(p_X)$ or $y^n \notin T_\delta^n(p_Y)$ then $(x^n, y^n) \notin T_\delta^n(p_{XY})$. Since we are given that $x^n \notin T_\delta^n(p_X)$ then for all $y^n \in \mathcal{Y}^n$, $(x^n, y^n) \notin T_\delta^n(p_{XY})$ because otherwise by Problem 3.(5) $x^n \in T_\delta^n(p_X)$ which is false. Hence for all $y^n \in \mathcal{Y}^n$, $(x^n, y^n) \notin T_\delta^n(p_{XY})$ therefore $T_\delta^n(p_{XY}|x^n) = \emptyset$
5. $p_{Y|X}^n(y^n|x^n) = \prod_{i=1}^n p_{Y|X}(y_i|x_i) = \frac{p_{XY}^n(x, y)}{p_X^n(x^n)}$. Now from Problem 3.(3) we have

$$2^{-n[H(X, Y) + \delta]} \leq p_{XY}^n(x^n, y^n) \leq 2^{-n[H(X, Y) - \delta]}$$

and from Problem 2.(3) we get

$$x^n \in T_\delta^n(p_X) \implies 2^{-n(H(X) + \delta)} \leq \prod_{i=1}^n p_X(x_i) \leq 2^{-n(H(X) - \delta)}$$

Therefore we have

$$\frac{2^{-n[H(X, Y) + \delta]}}{2^{-n(H(X) - \delta)}} \leq \frac{p_{XY}^n(x, y)}{p_X^n(x^n)} \leq \frac{2^{-n[H(X, Y) - \delta]}}{2^{-n(H(X) + \delta)}}$$

Now

$$\frac{2^{-n[H(X,Y)+\delta]}}{2^{-n(H(X)-\delta)}} = 2^{-n((H(X,Y)+\delta)-(H(X)-\delta))} = 2^{-n(H(X,Y)-H(X)+2\delta)} = 2^{-n(H(Y|X)+2\delta)}$$

Similarly

$$\frac{2^{-n[H(X,Y)-\delta]}}{2^{-n(H(X)+\delta)}} = 2^{-n((H(X,Y)-\delta)-(H(X)+\delta))} = 2^{-n(H(X,Y)-H(X)-2\delta)} = 2^{-n(H(Y|X)+2\delta)}$$

Therefore we get

$$2^{-n(H(Y|X)+2\delta)} \leq p_{Y|X}^n(y^n|x^n) \leq 2^{-n(H(Y|X)+2\delta)}$$

6.

7. Using part (4) we have

$$1 \geq \sum_{y^n \in T_{2\delta}^n(p_{XY}|x^n)} p_{Y|X}^n(y|x) \geq \sum_{y^n \in T_{2\delta}^n(p_{XY}|x^n)} 2^{-n(H(Y|X)+2\delta)} = |T_{2\delta}^n(p_{XY}|x^n)| 2^{-n(H(Y|X)+2\delta)}$$

Therefore we get

$$|T_{2\delta}^n(p_{XY}|x^n)| \leq 2^{n(H(Y|X)+2\delta)}$$

Now in part (5) we proved that

$$Pr[y^n \in T_{2\delta}^n(p_{XY}|x^n)] \geq 1 - 2|\mathcal{X}||\mathcal{Y}| \exp \left[-\frac{2n\delta^2}{(\log |\mathcal{X}||\mathcal{Y}|)^2} \eta_{p_{Y|X}} \right]$$

Also

$$\begin{aligned} Pr[y^n \in T_{2\delta}^n(p_{XY}|x^n)] &= \sum_{y^n \in T_{2\delta}^n(p_{XY}|x^n)} p_{Y|X}^n(y^n|x^n) \\ &\leq \sum_{y^n \in T_{2\delta}^n(p_{XY}|x^n)} 2^{-n(H(Y|X)+2\delta)} = |T_{2\delta}^n(p_{XY}|x^n)| 2^{-n(H(Y|X)+2\delta)} \end{aligned}$$

Hence we get

$$\begin{aligned} |T_{2\delta}^n(p_{XY}|x^n)| 2^{-n(H(Y|X)+2\delta)} &\geq 1 - 2|\mathcal{X}||\mathcal{Y}| \exp \left[-\frac{2n\delta^2}{(\log |\mathcal{X}||\mathcal{Y}|)^2} \eta_{p_{Y|X}} \right] \\ \implies |T_{2\delta}^n(p_{XY}|x^n)| &\geq \left[1 - 2|\mathcal{X}||\mathcal{Y}| \exp \left(-\frac{2n\delta^2}{(\log |\mathcal{X}||\mathcal{Y}|)^2} \eta_{p_{Y|X}} \right) \right] 2^{n(H(Y|X)+2\delta)} \end{aligned}$$

Therefore we finally get

$$\left[1 - 2|\mathcal{X}||\mathcal{Y}| \exp \left(-\frac{2n\delta^2}{(\log |\mathcal{X}||\mathcal{Y}|)^2} \eta_{p_{Y|X}} \right) \right] 2^{n(H(Y|X)+2\delta)} 2 \leq |T_{2\delta}^n(p_{XY}|x^n)| \leq 2^{n(H(Y|X)+2\delta)}$$

□

Given that $x^n \in T_\delta^n(p_X(x))$ and $T_\delta^n(p_{XY}) = \{y^n \in \mathcal{Y}^n \mid (x^n, y^n) \in T_\delta^n(p_{XY}(x, y))\}$. Now for any $a \in \mathcal{X}$ let $S_a = \{i \in [n] \mid x_i = a\}$. Then naturally $|S_a| = N(a|x^n)$. Now consider the projection of y^n on to S_a . Let's denote it by y_a^n . Now

$$\sum_{i=1}^n \mathbb{1}_{(x_i, y_i) = (a, b)} = \sum_{i=1}^{N(a|x^n)} \mathbb{1}_{y_{a_i}^n = b}$$

But here

$$\Pr \left[\mathbb{1}_{y_{a_i}^n = b} = 1 \right] = \Pr[Y = b | X = a] = p_{Y|X}(b|a)$$

Since $(x^n, y^n) \in T_\delta^n(p_{XY})$ we have $y^n|_{S_a} \in T_\delta^{N(a|x^n)}(p_{Y|X=a})$ where $P_{Y|X=a}(y)$ is the probability distribution over $y \in \mathcal{Y}$. Ans since a is any arbitrary element of \mathcal{X} , $y^n|_{S_a} \in T_\delta^{N(a|x^n)}(p_{Y|X=a})$ should be true for all $a \in \mathcal{X}$. Hence

$$y^n \in T_\delta^n(p_{XY}|x^n) \iff \forall a \in \mathcal{X}, y_a^{N(a|x^n)} \in T_\delta^{N(a|x^n)}(p_{Y|X=a})$$

where when going from left to right we mean $y_a^{N(a|x^n)} := y^n|_{S_a}$ and when going from right to left y^n is constructed by putting the elements in each $y_a^{N(a|x^n)}$ in their corresponding positions according to x^n .

Therefore

$$y^n \notin T_\delta^n(p_{XY}|x^n) \iff \exists a \in \mathcal{X}, y_a^{N(a|x^n)} \notin T_\delta^{N(a|x^n)}(p_{Y|X=a})$$

Now

$$T_\delta^{N(a|x^n)}(p_{Y|X=a}) = \left\{ \left| \frac{N(b|y_a^{N(a|x^n)})}{N(a|x^n)} - p_{Y|X=a}(b) \right| \leq \frac{\delta p_{Y|X=a}}{\log |\mathcal{Y}|} \right\}$$

By Hoeffding inequality take the $\mathbb{1}_{y_i=b}$ for $i \in [N(a|x^n)]$ as random variables then

$$\Pr \left[\left| \frac{1}{N(a|x^n)} \sum_{i=1}^{N(a|x^n)} \mathbb{1}_{y_i=b} - p_{Y|X=a}(b) \right| > \frac{\delta p_{Y|X=a}}{\log |\mathcal{Y}|} \right] \leq 2 \exp \left[-2N(a|x^n) \frac{\delta^2 p_{Y|X=a}^2(b)}{\log^2 |\mathcal{Y}|} \right]$$