



# NUS

National University  
of Singapore

Name : LUO ZIJIAN

Matric.No: A0224725H

MUSNET: E0572844

Subject: Information Theory

Assignment: Homework Six

Date: October 10<sup>th</sup>

Prof: Marco Tomamichel

## Exercise 6.1

Proof:

As we know,  $P_Y(y) = \sum_{x \in \mathcal{X}} T(y|x) \cdot P(x)$ ,  $Q_Y(y) = \sum_{x \in \mathcal{X}} T(y|x) \cdot Q(x)$

And for the definition of total variation distance, we can show that

$$\delta_{\text{td}}(P_X, Q_X) = \frac{1}{2} \|P_X - Q_X\| = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|$$

$$\delta_{\text{td}}(P_Y, Q_Y) = \frac{1}{2} \|P_Y - Q_Y\| = \frac{1}{2} \sum_{y \in \mathcal{Y}} |P(y) - Q(y)| = \frac{1}{2} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} [P(x) - Q(x)] \cdot T(y|x)$$

With the definition of data processing inequality in KL

$$D(P_X \| Q_X) \geq D(P_Y \| Q_Y)$$

And then

$$\delta_{\text{td}}(P_X, Q_X) - \delta_{\text{td}}(P_Y, Q_Y) = \frac{1}{2} \sum_{x \in \mathcal{X}} [P(x) - Q(x)] \left[ 1 - \sum_{y \in \mathcal{Y}} T(y|x) \right] \quad (1 \geq \sum_{y \in \mathcal{Y}} T(y|x))$$

$$\geq 0.$$

Finally, it is desired.

## Exercise 6.2

Proof: As for the definition of empirical set,

we know the empirical distribution:  $f_{X^n}(x) := \frac{1}{n} |\{i \in \{1, \dots, n\} : x_i = x\}|$

And use the expression of total variation distance, we can rewrite empirical typical set like that

$$\begin{aligned} \delta_{\text{td}}(f_{X^n}, P(x)) &\leq \epsilon \\ &= \frac{1}{2} \|f_{X^n} - P(x)\| \leq \epsilon \\ &= \frac{1}{2} \left\| \frac{1}{n} |\{i \in \{1, \dots, n\} : x_i = x\}| - P(x) \right\| \leq \epsilon \end{aligned}$$

And then, we know  $x_1, x_2, \dots, x_n$  are a sequence of i.i.d. source.

$$\frac{1}{n} |\{i \in \{1, \dots, n\} : x_i = x\}| \rightarrow u \quad \text{for } n \rightarrow \infty$$

In this case  $u = P(x)$ .





And we use the ~~large~~ law of large number ~~to~~, which shows that

$$Pr \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = u \right\} = 1 \quad (\text{almost sure convergence})$$

In a word, the empirical distribution of a sequence should be "close" to the prior probability distribution.

$$Pr \left[ P \left[ x^n \in A_{\text{emp}, \varepsilon}^{(n)} \mid P(x) \right] = 1 \right]$$

### Exercise 6.3.

a) Symmetric error probabilities:  ~~$\varepsilon^*$~~   ~~$f(P, Q, S)$~~

$$C(P, Q, S) = - \min_{0 \leq \lambda_1 \leq 1, 0 \leq \lambda_2 \leq 1} \log \sum_{x \in X} P(x)^{\lambda_1} Q(x)^{\lambda_2} S(x)^{1-\lambda_1-\lambda_2}$$

especially when  $\lambda_1 = \lambda_2 = \frac{1}{3}$ , ~~we~~ we can get minimum value.

$$C(P, Q, S) = -\frac{1}{3}$$

$$\varepsilon_{\text{sym}, 1}^* = 2^{-n \cdot C(P, Q, S)} = 2^{-\frac{n}{3}} = 2^{-\frac{1}{3}} = \frac{1}{8}$$

(b). In ternary hypothesis testing problem, ~~we can~~ ~~to~~

especially when  $\lambda_1 = \lambda_2 = \frac{1}{3}$ , all three have equal priors

That is  $Q(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

~~Therefore~~ Therefore, the minimal error probability:  $\frac{1}{8}$

In a word,  $Q$  is the optimal test.

$$D(P \parallel Q) = \sum_i p_i \log \frac{p_i}{q_i} = \sum_i \frac{1}{3} \log \frac{32}{27}$$

$$D(P \parallel S) = \sum_i s_i \log \frac{s_i}{p_i} = 1$$



# Exercise 6.4

	$x_1$	$x_2$	$x_3$	$x_4$	$x_1 \oplus x_2$	$x_3 \oplus x_4$	$x_1 \oplus x_3$	$x_2 \oplus x_4$
a).	0	0	0	0	0	0	0	0
	0	0	0	1	0	1	0	1
	0	0	0	1	0	1	1	0
	0	0	1	0	0	0	1	1
	0	0	1	1	0	0	0	1
	0	1	0	0	1	1	0	0
	0	1	0	1	1	1	1	1
	0	1	1	0	1	0	1	0
	0	1	1	1	1	0	1	0
	1	0	0	0	1	1	1	1
	1	0	0	1	1	1	1	0
	1	0	1	0	1	1	0	1
	1	0	1	1	1	0	1	1
	1	1	0	0	0	0	1	0
	1	1	0	1	0	1	0	1
	1	1	1	0	0	1	0	1
	1	1	1	1	0	0	0	0

minimal distance : 3

detect errors : 2

correct errors : 1

(b) ~~Certainly, it is linear code.~~

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Obviously, it is linear code

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$





(c) Using the hamming bound,

it is not perfect because it can make one more redundancy, which means the most perfect code should be 7 bits

(d) Dual code

