Exercise 7.1

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

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

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

With the definition of data processing inequality in KL

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

And then

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

$\geq 0.$

Finally, it is desired.

Exercise 7.2

Proof: As for the definition of empirical set,

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

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

$\delta_{tvd}(f_{x^n}, P(x)) \leq \epsilon$

$= \frac{1}{2}\| f_{x^n} - P(x) \| \leq \epsilon$

$= \frac{1}{2}\left\| \frac{1}{n}\left|\{i \in \{1, \cdots n\} : x_i = x\}\right| - P(x) \right\| \leq \epsilon$

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

$\left|\{i \in \{1, \cdots n\} : x_i = x\}\right| \to u$ for $n \to \infty$

In this case $u = n \cdot P(x)$.

FALCON

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

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

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

$$\Pr_{n\to\infty} P\left[x^n \in A^{(n)}_{emp,\varepsilon} P(x)\right] = 1$$

Exercise 7.3

a). Symmetric error probabilities:

$$C(P,Q,S) = -\min_{0\le\lambda_1\le1,\ 0\le\lambda_2\le1} \log \sum_{x\in X} P(x)^{\lambda_1} Q(x)^{\lambda_2} S(x)^{1-\lambda_1-\lambda_2}$$

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

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

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

(b). In ternary hypothesis testing problem, especially when $\lambda_1 = \lambda_2 = \frac{1}{3}$, all three have equal priors

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

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

In a word, Q is the optimal test.

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

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

Exercise 7.4

a).

x_1	x_2	x_3	x_4	$x_1 \oplus x_2$	$x_3 \oplus x_4$	$x_1 \oplus x_3$	$x_2 \oplus x_4$
0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1
0	0	1	0	0	1	1	0
0	0	1	1	0	0	1	1
0	1	0	0	1	0	0	1
0	1	0	1	1	1	0	0
0	1	1	0	1	1	1	1
0	1	1	1	1	0	1	0
1	0	0	0	1	0	1	0
1	0	0	1	1	1	1	1
1	0	1	0	1	1	0	0
1	0	1	1	1	0	0	1
1	1	0	0	0	0	1	1
1	1	0	1	0	1	1	0
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 redundcy, which means the most perfect code should be 7 bits

(d) Dual code