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

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

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

With the definition of data processing inequality in KL

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

And then

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

$\geq 0.$

Finally, it is desired.

Exercise 6.2

Proof: As for the definition of empirical set,

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

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

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

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

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

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

$|\{i\in\{1,\ldots,n\} : x_i = x\}| \rightarrow u$ for $n\rightarrow\infty$

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

FALCON

扫描全能王 创建

And we use the ~~large~~ law of large number ~~s~~, 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 6.3.

a). Symmetric error probabilities : ~~ε^*_{sym}~~ ~~(P, Q)~~

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

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, ~~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.

(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 6.4

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) ~~Certainty, it is linear code.~~

~~G: [1 0 0 0; 0 1 0 0; 0 0 1 0; 0 0 0 1]~~

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 redun cy, which means the most perfect code should be 7 bits

(d) Dual code

扫描全能王 创建