

Homework 3

YOUR NAME

October 26, 2020

- **Acknowledgments:** For Problem 3.3(b), I refer to the lecture notes of Madhur Tulsiani, University of Chicago <https://ttic.uchicago.edu/~madhurt/courses/infotheory2014/15.pdf>. For Problem 3.6(b)-(d), I refer to the essay **Detection with Distributed Sensors** by Robert R. Tenney <https://ieeexplore.ieee.org/abstract/document/4102537>
- **Collaborators:** I finish this template by myself.
- *I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.*

Hanmo Chen

3.1. (a) First we have,

$$\begin{aligned} \sum_{x \in \mathcal{X}} P_y(x) D(P_{y|x=x} \| P_{y|x=x_0}) &= \sum_{x \in \mathcal{X}} P_y(x) \left(\sum_{y \in \mathcal{Y}} P_{y|x=x}(y) \log \frac{P_{y|x=x}(y)}{P_{y|x=x_0}(y)} \right) \\ &= \sum_{x \in \mathcal{X}} P_x(x) \left(\sum_{y \in \mathcal{Y}} P_{y|x=x}(y) \log \frac{P_{x,y}(x, y)}{P_x(x) P_{y|x=x_0}(y)} \right) \\ &= \sum_{x \in \mathcal{X}} P_x(x) \left[\sum_{y \in \mathcal{Y}} P_{y|x=x}(y) \left(\log \frac{P_{x,y}(x, y)}{P_x(x) P_y(y)} + \log \frac{P_y(y)}{P_{y|x=x_0}(y)} \right) \right] \\ &= \sum_{x, y \in \mathcal{X} \times \mathcal{Y}} P_x(x) P_{y|x=x}(y) \log \frac{P_{x,y}(x, y)}{P_x(x) P_y(y)} \\ &\quad + \sum_{x, y \in \mathcal{X} \times \mathcal{Y}} P_x(x) P_{y|x=x}(y) \log \frac{P_y(y)}{P_{y|x=x_0}(y)} \end{aligned} \tag{1}$$

And because $P_x(x) P_{y|x=x}(y) = P_{x,y}(x, y)$ and $\log \frac{P_y(y)}{P_{y|x=x_0}(y)}$ is independent of x , we have,

$$\begin{aligned} \sum_{x \in \mathcal{X}} P_x(x) D(P_{y|x=x} \| P_{y|x=x_0}) &= I(X; Y) + \sum_{y \in \mathcal{Y}} P_y(y) \log \frac{P_y(y)}{P_{y|x=x_0}(y)} \\ &= I(X; Y) + D(P_y \| P_{y|x=x_0}) \end{aligned} \tag{2}$$

That is,

$$I(X; Y) = \sum_{x \in \mathcal{X}} P_x(x) D(P_{y|x=x} \| P_{y|x=x_0}) - D(P_y \| P_{y|x=x_0}) \quad (3)$$

(b) From the last formula and $D(P_y \| P_{y|x=x_0}) \geq 0$ we know that $\forall x_0$

$$I(X; Y) \leq \sum_{x \in \mathcal{X}} P_x(x) D(P_{y|x=x} \| P_{y|x=x_0}) \quad (4)$$

And $\sum_{x \in \mathcal{X}} P_x(x) = 1$, thus

$$\sum_{x \in \mathcal{X}} P_x(x) D(P_{y|x=x} \| P_{y|x=x_0}) \leq \sup_{x \in \mathcal{X}} D(P_{y|x=x} \| P_{y|x=x_0}) \quad (5)$$

So

$$I(X; Y) \leq \sup_{x, x_0 \in \mathcal{X}} D(P_{y|x=x} \| P_{y|x=x_0}) \quad (6)$$

3.2. (a) First we prove that $H(x; y; z) \leq H(x; y) + H(x; z) - H(x)$

$$\begin{aligned} H(x; y; z) &= H(y; z | x) - H(x) \\ &\leq H(y | x) + H(z | x) - H(x) = H(x; y) + H(x; z) - H(x) \end{aligned} \quad (7)$$

Symmetrically, $H(x; y; z) \leq H(x; y) + H(y; z) - H(y)$, thus

$$\begin{aligned} 2H(x; y; z) &\leq 2H(x; y) + H(x; z) + H(y; z) - H(x) - H(y) \\ &\leq H(x; y) + H(x; z) + H(y; z) \end{aligned} \quad (8)$$

(b) Place n points in \mathbb{R}^3 arbitrarily. Let (x, y, z) be the random variables uniformly distributed among those points, so $H(x, y, z) = \log(n)$.

$$2H(x; y; z) \leq H(x; y) + H(x; z) + H(y; z) \leq \log |\mathcal{X} \times \mathcal{Y}| + \log |\mathcal{Y} \times \mathcal{Z}| + \log |\mathcal{X} \times \mathcal{Z}| = \log(n_1 n_2 n_3) \quad (9)$$

Thus

$$n_1 n_2 n_3 \geq n^2 \quad (10)$$

3.3. (a) To simplify, we use the natural logarithm. denote $f(p, q) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q} - 2(p-q)^2$.

$$\begin{aligned} \frac{\partial f}{\partial q} &= \frac{p}{q} - \frac{1-p}{1-q} - 4(p-q) = \frac{p-q}{q(1-q)} - 4(p-q) \\ &= (p-q) \left[\frac{1}{q(1-q)} - 4 \right] \end{aligned} \quad (11)$$

For fixed p , $\frac{\partial f}{\partial q} \leq 0$ when $q < p$ and $\frac{\partial f}{\partial q} \geq 0$ when $q > p$. So to minimize $f(p, q)$, $q_{\min} = p$ and $f(p, p) = 0$, therefore

$$p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q} \geq 2(p-q)^2 \quad (12)$$

(b) Also use the natural logarithm. We need to prove

$$\text{TV}(P, Q) \triangleq \sup_{E \in \mathcal{F}} (P(E) - Q(E)) \leq \sqrt{\frac{1}{2} D(P \| Q)} \quad (13)$$

Consider the event $E^* = \{w | P(w) \geq Q(w)\}$, so that $\text{TV}(P, Q) \triangleq \sup_{E \in \mathcal{F}} (P(E) - Q(E)) = P(E^*) - Q(E^*)$. And $P(E^*) + P(E^{*c}) = Q(E^*) + Q(E^{*c}) = 1$, so $|P(E^*) - Q(E^*)| = |P(E^{*c}) - Q(E^{*c})|$, which means,

$$\begin{aligned} \text{TV}(P, Q) &= P(E^*) - Q(E^*) \\ &= \frac{1}{2} (|P(E^*) - Q(E^*)| + |P(E^{*c}) - Q(E^{*c})|) \\ &= \frac{1}{2} \sum_{w \in \Omega} |P(w) - Q(w)| = \frac{1}{2} \|P - Q\|_1 \end{aligned} \quad (14)$$

Updating Equation 13 with Equation 14, we just need to prove

$$\frac{1}{2} \|P - Q\|_1^2 \leq D(P \| Q) \quad (15)$$

Let X be the indicator random variable of event E^* , that is

$$X(w) = \mathbf{I}(E^*) = \begin{cases} 1, & \text{if } P(w) \geq Q(w) \\ 0, & \text{if } P(w) < Q(w) \end{cases} \quad (16)$$

So X is a bernoulli random variable with probability $P(E^*), Q(E^*)$, which we denote as the bernoulli distribution P_X, Q_X with probability p, q .

Using the conclusion from (a),

$$\frac{1}{2} \|P - Q\|_1^2 = \frac{1}{2} (|p - q| + |(1-p) - (1-q)|)^2 = 2(p-q)^2 \leq D(P_X \| Q_X) \quad (17)$$

And use the data processing inequality $D(P_X \| Q_X) \leq D(P \| Q)$, so we prove that

$$\text{TV}(P, Q) \leq \sqrt{\frac{1}{2} D(P \| Q)} \quad (18)$$

3.4. (a) Decision rule

$$\begin{aligned}
\hat{H}(y) &= \arg \max_{H_i, i=0,1} p_{H|Y}(H_i|y) \\
&= \arg \max_{H_i, i=0,1} \frac{p_{Y|H}(y|H_i)p(H_i)}{\sum_{i=0,1} p_{Y|H}(y|H_i)p(H_i)} \\
&= \arg \max_{H_i, i=0,1} p_{Y|H}(y|H_i)p(H_i)
\end{aligned} \tag{19}$$

- For H_0 , $p_{Y|H}(y|H_0)p(H_0) = p, 0 \leq y \leq 1$
- For H_1 , $p_{Y|H}(y|H_1)p(H_1) = 2y(1-p), 0 \leq y \leq 1$

So if $p < \frac{2}{3}$

$$\hat{H}(y) = \arg \max_{H_i, i=0,1} p_{Y|H}(y|H_i)p(H_i) = \begin{cases} H_0, & y \leq \frac{p}{2(1-p)} \\ H_1, & y > \frac{p}{2(1-p)} \end{cases} \tag{20}$$

If $p \geq \frac{2}{3}$, $\hat{H}(y) = H_0$.

(b) The Likelihood Ratio is,

$$\text{LRT}(y) = \frac{p_{Y|H}(y|H_0)}{p_{Y|H}(y|H_1)} = \frac{1}{2y} \geq \lambda \tag{21}$$

For P_D , let $c = \frac{1}{2\lambda}, \lambda \in [\frac{1}{2}, \infty)$

$$P_D \triangleq \mathbb{P}(\hat{H} = H_1 | H = H_1) = \mathbb{P}(y > c | H = H_1) = 1 - c^2 \tag{22}$$

For P_F ,

$$P_F \triangleq \mathbb{P}(\hat{H} = H_1 | H = H_0) = \mathbb{P}(y > c | H = H_0) = 1 - c \tag{23}$$

Replace c with $1 - P_F$ in equation 22, we find the operating characteristic of LRT as

$$P_D = (2 - P_F)P_F \tag{24}$$

- (c) i. Because $P_D = (2 - P_F)P_F$, let $P_D \geq (1 + \epsilon)P_F$, we have $P_F \leq (1 - \epsilon)$. And P_D is a monotonically increasing function of P_F in $[0, 1]$, so $P_D^{\max}(\epsilon) = 1 - \epsilon^2$
- ii. $P_D^{\max}(\epsilon) = 1 - \epsilon^2 > 0$ so $\epsilon \in (0, 1)$
- iii. In the decision rule from part (a), $P_F = 1 - \frac{p}{2(1-p)} \leq 1 - \epsilon$, that is $p \geq \frac{2\epsilon}{1+2\epsilon}$

3.5. (a) i. The cost is

$$\mathbb{E}[c(f(\underline{y}), H)] = \sum_{i=1}^3 P_{\underline{y}}(\underline{y}) \mathbf{I}(f(\underline{y}) = H_i) \left(\sum_{j=1}^3 C_{ij} \pi_j(\underline{y}) \right) \quad (25)$$

So

$$\hat{H} = f(\underline{y}) = \arg \min_i \sum_{j=1}^3 C_{ij} \pi_j(\underline{y}) \quad (26)$$

Specifically,

$$\begin{cases} l_1(\underline{y}) = \pi_2(\underline{y}) + 2\pi_3(\underline{y}) \\ l_2(\underline{y}) = \pi_1(\underline{y}) + 2\pi_3(\underline{y}) \\ l_3(\underline{y}) = 2\pi_1(\underline{y}) + 2\pi_2(\underline{y}) \end{cases} \quad (27)$$

$\alpha)$ $\hat{H} = H_1$: let $l_1(\underline{y}) < l_2(\underline{y})$ and $l_1(\underline{y}) < l_3(\underline{y})$ we have
 $\pi_1(\underline{y}) > \pi_2(\underline{y})$ and $2\pi_3(\underline{y}) < 2\pi_1(\underline{y}) + \pi_2(\underline{y})$

$\beta)$ $\hat{H} = H_2$: let $l_2(\underline{y}) < l_1(\underline{y})$ and $l_2(\underline{y}) < l_3(\underline{y})$ we have
 $\pi_1(\underline{y}) < \pi_2(\underline{y})$ and $2\pi_3(\underline{y}) < \pi_1(\underline{y}) + 2\pi_2(\underline{y})$

$\gamma)$ $\hat{H} = H_3$: let $l_3(\underline{y}) < l_2(\underline{y})$ and $l_3(\underline{y}) < l_1(\underline{y})$ we have
 $2\pi_3(\underline{y}) > 2\pi_1(\underline{y}) + \pi_2(\underline{y})$ and $2\pi_3(\underline{y}) > \pi_1(\underline{y}) + 2\pi_2(\underline{y})$

So the optimum decision rule is,

$$\hat{H} = f(\underline{y}) = \begin{cases} H_1, & \pi_1(\underline{y}) > \pi_2(\underline{y}) \text{ and } 2\pi_3(\underline{y}) < 2\pi_1(\underline{y}) + \pi_2(\underline{y}) \\ H_2, & \pi_1(\underline{y}) < \pi_2(\underline{y}) \text{ and } 2\pi_3(\underline{y}) < \pi_1(\underline{y}) + 2\pi_2(\underline{y}) \\ H_3, & 2\pi_3(\underline{y}) > 2\pi_1(\underline{y}) + \pi_2(\underline{y}) \text{ and } 2\pi_3(\underline{y}) > \pi_1(\underline{y}) + 2\pi_2(\underline{y}) \end{cases} \quad (28)$$

ii. Using $\pi_1 + \pi_2 + \pi_3 = 1$

$$\hat{H} = f(\underline{y}) = \begin{cases} H_1, & \pi_1 > \pi_2 \text{ and } 4\pi_1 + 3\pi_2 > 2 \\ H_2, & \pi_2 > \pi_1 \text{ and } 3\pi_1 + 4\pi_2 > 2 \\ H_3, & 4\pi_1 + 3\pi_2 < 2 \text{ and } 3\pi_1 + 4\pi_2 < 2 \end{cases} \quad (29)$$

And the decision regions in the (π_1, π_2) plane is,

(b) When $C_{ij} = 1 - \delta_{ij}$, the optimum decision rule is

$$\hat{H} = f(\underline{y}) = \arg \min_i \sum_{j=1}^3 C_{ij} \pi_j(\underline{y}) = \arg \max_i \pi_i(\underline{y}) \quad (30)$$

Consider $L_i(\underline{y}) = \frac{p_{\underline{y}|\mathbf{H}}(\underline{y}|H_i)}{p_{\underline{y}|\mathbf{H}}(\underline{y}|H_1)}$ for $i = 2, 3$, because three hypotheses are equally likely a priori,

$$L_i(\underline{y}) = \frac{p_{\underline{y}|\mathbf{H}}(\underline{y} | H_i)}{p_{\underline{y}|\mathbf{H}}(\underline{y} | H_1)} = \frac{\pi_i(\underline{y})}{\pi_1(\underline{y})} \quad (31)$$

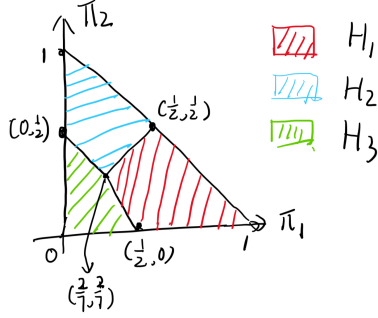


Figure 1: Decision regions in the (π_1, π_2) plane

For consistency, denote $L_1(\underline{y}) = 0$, so the optimum decision can be expressed in terms of $L_i(\underline{y})$ as,

$$\hat{H} = \arg \max_i \pi_i(\underline{y}) = \arg \max_i L_i(\underline{y}) \quad (32)$$

To calculate $L_i(\underline{y})$,

$$L_i(\underline{y}) = \frac{\exp(-y_1^2 - (y_i - 1)^2)}{\exp(-y_i^2 - (y_1 - 1)^2)} = \exp(2(y_i - y_1)), i = 2, 3 \quad (33)$$

So the optimum decision rule can be specified in terms of the pair of sufficient statistics

$$\begin{aligned} \ell_2(\underline{y}) &= y_2 - y_1 \\ \ell_3(\underline{y}) &= y_3 - y_1 \end{aligned} \quad (34)$$

And the optimum decision rule is,

$$\hat{H} = f(\underline{y}) = \begin{cases} H_1, \ell_2(\underline{y}) < 0 \text{ and } \ell_2(\underline{y}) < 0 \\ H_2, \ell_2(\underline{y}) > 0 \text{ and } \ell_2(\underline{y}) > \ell_3(\underline{y}) \\ H_3, \ell_3(\underline{y}) > 0 \text{ and } \ell_3(\underline{y}) > \ell_2(\underline{y}) \end{cases} \quad (35)$$

- 3.6. (a) At the local sensor, $P(y_i > 0|x = 1) = P(y_i < 0|x = -1) = 0.75$,
 $P(y_i < 0|x = 1) = P(y_i > 0|x = -1) = 0.25$.

Thus, the conditional distribution of \hat{x}_i given x is,

$$\begin{aligned} P(\hat{x}_i|x = 1) &= 0.75\mathbf{I}(\hat{x}_i=1)0.25\mathbf{I}(\hat{x}_i=-1) \\ P(\hat{x}_i|x = -1) &= 0.75\mathbf{I}(\hat{x}_i=-1)0.25\mathbf{I}(\hat{x}_i=1) \end{aligned} \quad (36)$$

At the fusion center,

$$\begin{aligned}
P(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n | x = 1) &= \prod_{i=1}^n P(\hat{x}_i | x = 1) \\
&= 0.75^{\sum_{i=1}^n \mathbf{I}(\hat{x}_i = 1)} 0.25^{\sum_{i=1}^n \mathbf{I}(\hat{x}_i = -1)} \\
P(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n | x = -1) &= \prod_{i=1}^n P(\hat{x}_i | x = -1) \\
&= 0.75^{\sum_{i=1}^n \mathbf{I}(\hat{x}_i = -1)} 0.25^{\sum_{i=1}^n \mathbf{I}(\hat{x}_i = 1)}
\end{aligned} \tag{37}$$

Denote $m = \sum_{i=1}^n \mathbf{I}(\hat{x}_i = 1)$ as the number of 1-s in $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$, and because of equal probability priori,

$$\begin{aligned}
\frac{P(x = 1 | \hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)}{P(x = -1 | \hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)} &= \frac{P(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n | x = 1)}{P(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n | x = -1)} \\
&= 3^{2m-n}
\end{aligned} \tag{38}$$

So the minimum probability of error decision rule is a majority-voting rule as follows,

$$\hat{x}(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = \begin{cases} 1, & m \geq \frac{n}{2} \\ -1, & m < \frac{n}{2} \end{cases} \tag{39}$$

where m denotes the number of one-s in $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$.

- (b) Suppose the local decision rule can be expressed as the conditional distriburion $p(\hat{x}_i | y_i)$ of \hat{x}_i given $y_i, i = 1, 2$.

Part 1

First we prove that the decision rule is a deterministic rule, that is, given y_i , \hat{x}_i is determined.

The expected joint cost is

$$\begin{aligned}
\mathbb{E}[C(\hat{x}_1, \hat{x}_2, x)] &= \sum_{\hat{x}_1, \hat{x}_2, x} C(\hat{x}_1, \hat{x}_2, x) p(\hat{x}_1, \hat{x}_2, x) \\
&= \sum_{\hat{x}_1, \hat{x}_2, y_1, y_2, x} C(\hat{x}_1, \hat{x}_2, x) p(\hat{x}_1, \hat{x}_2, y_1, y_2, x) \\
&= \sum_{\hat{x}_1, \hat{x}_2, y_1, y_2, x} C(\hat{x}_1, \hat{x}_2, x) p(\hat{x}_1 | y_1) p(\hat{x}_2 | y_2) p(y_1 | x) p(y_2 | x) p(x) \\
&= \sum_{\hat{x}_2, y_1, y_2, x} \lambda(\hat{x}_2, y_1, y_2, x) [C(1, \hat{x}_2, x) p(\hat{x}_1 = 1 | y_1) + C(-1, \hat{x}_2, x) p(\hat{x}_1 = -1 | y_1)] \\
&\quad \text{where } \lambda(\hat{x}_2, y_1, y_2, x) = p(\hat{x}_2 | y_2) p(y_1 | x) p(y_2 | x) p(x) \\
&\quad \text{(Expanding the sum over } \hat{x}_1)
\end{aligned} \tag{40}$$

Assume y_1 is given, because $p(\hat{x}_1 = -1 | y_1) = 1 - p(\hat{x}_1 = 1 | y_1)$,

$$\begin{aligned} \mathbb{E}[C(\hat{x}_1, \hat{x}_2, x)|y_1] &= \sum_{\hat{x}_2, x} p_{\hat{x}_2, x}(\hat{x}_2, x) p_{y_1|x}(y_1|x) \times \\ &\quad [(C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)) p(\hat{x}_1 = 1|y_1) + C(-1, \hat{x}_2, x)] \end{aligned} \quad (41)$$

To minimize $\mathbb{E}[C(\hat{x}_1, \hat{x}_2, x)|\hat{x}_2, y_1]$,

$$p(\hat{x}_1 = 1|y_1) = \begin{cases} 0, & \sum_{\hat{x}_2, x} p_{\hat{x}_2, x}(\hat{x}_2, x) p_{y_1|x}(y_1|x) [C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)] > 0 \\ 1, & \sum_{\hat{x}_2, x} p_{\hat{x}_2, x}(\hat{x}_2, x) p_{y_1|x}(y_1|x) [C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)] < 0 \end{cases} \quad (42)$$

which equals to,

$$\hat{x}_1(y_1) = \begin{cases} 1, & \sum_{\hat{x}_2, x} p_{\hat{x}_2, x}(\hat{x}_2, x) p_{y_1|x}(y_1|x) [C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)] < 0 \\ -1, & \sum_{\hat{x}_2, x} p_{\hat{x}_2, x}(\hat{x}_2, x) p_{y_1|x}(y_1|x) [C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)] > 0 \end{cases} \quad (43)$$

Part 2

In this part we will prove the decision rule in Equation 43 can be seen as a form of a likelihood ratio test .

Expanding the sum over x ,

$$\begin{aligned} &\sum_{\hat{x}_2, x} p_{\hat{x}_2, x}(\hat{x}_2, x) p_{y_1|x}(y_1|x) [C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)] \\ &= \sum_{\hat{x}_2} p_{\hat{x}_2, x}(\hat{x}_2, 1) p_{y_1|x}(y_1|1) [C(1, \hat{x}_2, 1) - C(-1, \hat{x}_2, 1)] \\ &\quad + \sum_{\hat{x}_2} p_{\hat{x}_2, x}(\hat{x}_2, -1) p_{y_1|x}(y_1|-1) [C(1, \hat{x}_2, -1) - C(-1, \hat{x}_2, -1)] \end{aligned} \quad (44)$$

Because the cost strictly increases with the number of errors made by the two sensors, so $C(1, \hat{x}_2, 1) < C(-1, \hat{x}_2, 1)$ and $C(1, \hat{x}_2, -1) > C(-1, \hat{x}_2, -1)$.

$$\sum_{\hat{x}_2, x} p_{\hat{x}_2, x}(\hat{x}_2, x) p_{y_1|x}(y_1|x) [C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)] < 0 \iff \frac{p_{y_1|x}(y_1|1)}{p_{y_1|x}(y_1|-1)} > \gamma_1 \quad (45)$$

where γ_1 depends on the rule $\hat{x}_2(\cdot)$,

$$\gamma_1 = \frac{\sum_{\hat{x}_2} p_{\hat{x}_2, x}(\hat{x}_2, -1) [C(1, \hat{x}_2, -1) - C(-1, \hat{x}_2, -1)]}{\sum_{\hat{x}_2} p_{\hat{x}_2, x}(\hat{x}_2, 1) [C(-1, \hat{x}_2, 1) - C(1, \hat{x}_2, 1)]} \quad (46)$$

And the decision rule for $\hat{x}_1(\cdot)$ is,

$$\hat{x}_1(y_1) = \begin{cases} 1, & \frac{p_{y_1|x}(y_1|1)}{p_{y_1|x}(y_1|-1)} > \gamma_1 \\ -1, & \frac{p_{y_1|x}(y_1|1)}{p_{y_1|x}(y_1|-1)} < \gamma_1 \end{cases} \quad (47)$$

(c) In the same way,

$$\hat{x}_2(y_2) = \begin{cases} 1, & \frac{p_{y_2|x}(y_2|1)}{p_{y_2|x}(y_2|-1)} > \gamma_2 \\ -1, & \frac{p_{y_2|x}(y_2|1)}{p_{y_2|x}(y_2|-1)} < \gamma_2 \end{cases} \quad (48)$$

where

$$\gamma_2 = \frac{\sum_{\hat{x}_1} p_{\hat{x}_1,x}(\hat{x}_1, -1) [C(\hat{x}_1, 1, -1) - C(\hat{x}_1, -1, -1)]}{\sum_{\hat{x}_1} p_{\hat{x}_1,x}(\hat{x}_1, 1) [C(\hat{x}_1, -1, 1) - C(\hat{x}_1, 1, 1)]} \quad (49)$$

(d) As the Equation 51 shows,

$$\begin{aligned} \gamma_1 &= \frac{p_{\hat{x}_2,x}(1, -1) [C(1, 1, -1) - C(-1, 1, -1)] + p_{\hat{x}_2,x}(-1, -1) [C(1, -1, -1) - C(-1, -1, -1)]}{p_{\hat{x}_2,x}(1, 1) [C(-1, 1, 1) - C(1, 1, 1)] + p_{\hat{x}_2,x}(-1, 1) [C(-1, -1, 1) - C(1, -1, 1)]} \\ &= \frac{p_{\hat{x}_2,x}(1, -1)(L-1) + p_{\hat{x}_2,x}(-1, -1)}{p_{\hat{x}_2,x}(1, 1) + p_{\hat{x}_2,x}(-1, 1)(L-1)} \end{aligned} \quad (50)$$

Because $p_{\hat{x}_2,x}(1, -1) + p_{\hat{x}_2,x}(-1, -1) = p_x(x = -1)$ and $p_{\hat{x}_2,x}(1, 1) + p_{\hat{x}_2,x}(-1, 1) = p_x(x = 1)$,

$$\gamma_1 = \frac{(L-2)p_{\hat{x}_2,x}(1, -1) + p_x(x = -1)}{(L-2)p_{\hat{x}_2,x}(-1, 1) + p_x(x = 1)} \quad (51)$$

So when $L = 2$, $\gamma_1 = \frac{p_x(x=-1)}{p_x(x=1)}$ does not depend on $\hat{x}_2(\cdot)$.

(e) **Some thoughts:**

It is also interesting when $L \neq 2$.

- When $1 < L < 2$, double errors are discounted relatively comparing to single error, which encourages local decisions to be more bold.
- When $L > 2$, double errors become more expensive. As L increases, to avoid double errors, \hat{x}_1 and \hat{x}_2 even may be always opposite!