Tsinghua-Berkeley Shenzhen Institute
Information Theory and Statistical Learning
Fall 2020

Homework 3

YOUR NAME October 23, 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/l5.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,

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

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

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

That is,

$$I(X;Y)=\sum_{x\in\mathcal{X}}P_{\mathrm{x}}(x)D\left(P_{\mathrm{y}|\mathrm{x}=x}\|P_{\mathrm{y}|\mathrm{x}=x_0}\right)-D(P_{\mathrm{y}}||P_{\mathrm{y}|\mathrm{x}=x_0}) \tag{3}$$

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

$$I(X;Y)\leqslant\sum_{x\in\mathcal{X}}P_{\mathrm{x}}(x)D\left(P_{\mathrm{y}|\mathrm{x}=x}\|P_{\mathrm{y}|\mathrm{x}=x_0}\right) \tag{4}$$

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

$$\sum_{x\in\mathcal{X}}P_{\mathrm{x}}(x)D\left(P_{\mathrm{y}|\mathrm{x}=x}\|P_{\mathrm{y}|\mathrm{x}=x_0}\right)\leqslant\sup_{x\in\mathcal{X}}D\left(P_{\mathrm{y}|\mathrm{x}=x}\|P_{\mathrm{y}|\mathrm{x}=x_0}\right) \tag{5}$$

So

$$I(X;Y)\leqslant\sup_{x,x_0\in\mathcal{X}}D\left(P_{\mathrm{y}|\mathrm{x}=x}\|P_{\mathrm{y}|\mathrm{x}=x_0}\right) \tag{6}$$

3.2. (a) First we prove that $H(\mathrm{x};\mathrm{y};\mathrm{z})\leqslant H(\mathrm{x};\mathrm{y})+H(\mathrm{x};\mathrm{z})-H(\mathrm{x})$

$$\begin{aligned}H(\mathrm{x};\mathrm{y};\mathrm{z})&=H(\mathrm{y};\mathrm{z}\mid\mathrm{x})-H(\mathrm{x})\\&\leqslant H(\mathrm{y}\mid\mathrm{x})+H(\mathrm{z}\mid\mathrm{x})-H(\mathrm{x})=H(\mathrm{x};\mathrm{y})+H(\mathrm{x};\mathrm{z})-H(\mathrm{x})\end{aligned} \tag{7}$$

Symmetrically, $H(\mathrm{x};\mathrm{y};\mathrm{z})\leqslant H(\mathrm{x};\mathrm{y})+H(\mathrm{y};\mathrm{z})-H(\mathrm{y})$, thus

$$\begin{aligned}2H(\mathrm{x};\mathrm{y};\mathrm{z})&\leqslant 2H(\mathrm{x};\mathrm{y})+H(\mathrm{x};\mathrm{z})+H(\mathrm{y};\mathrm{z})-H(\mathrm{x})-H(\mathrm{y})\\&\leqslant H(\mathrm{x};\mathrm{y})+H(\mathrm{x};\mathrm{z})+H(\mathrm{y};\mathrm{z})\end{aligned} \tag{8}$$

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

$$2H(\mathrm{x};\mathrm{y};\mathrm{z})\leqslant H(\mathrm{x};\mathrm{y})+H(\mathrm{x};\mathrm{z})+H(\mathrm{y};\mathrm{z})\leqslant\log|\mathcal{X}\times\mathcal{Y}|+\log|\mathcal{Y}\times\mathcal{Z}|+\log|\mathcal{X}\times\mathcal{Z}|=\log(n_1n_2n_3) \tag{9}$$

Thus

$$n_1n_2n_3\geqslant n^2 \tag{10}$$

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

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

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

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

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

$$\mathrm{TV}(P, Q) \triangleq \sup_{E \in \mathcal{F}} (P(E) - Q(E)) \leqslant \sqrt{2D(P\|Q)} \tag{13}$$

Consider the event $E^* = \{w | P(w) \geqslant Q(w)\}$, so that
$\mathrm{TV}(P, Q) \triangleq \sup_{E \in \mathcal{F}} (P(E) - Q(E)) = P(E^*) - Q(E^*)$
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,

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

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

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

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

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

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

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

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

$$\mathrm{TV}(P, Q) \leqslant \sqrt{2D(P\|Q)} \tag{18}$$

3.4. (a) Decision rule

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

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

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

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

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

(b) The Likelihood Ratio is,

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

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

$$P_{\mathrm{D}} \triangleq \mathbb{P}\left(\hat{\mathrm{H}} = H_1 \mid \mathrm{H} = H_1\right) = \mathbb{P}(y > c|H = H_1) = 1 - c^2 \tag{22}$$

For P_{F},

$$P_{\mathrm{F}} \triangleq \mathbb{P}\left(\hat{\mathrm{H}} = H_1 \mid \mathrm{H} = H_0\right) = \mathbb{P}(y > c|H = H_0) = 1 - c \tag{23}$$

Replace c with $1 - P_{\mathrm{F}}$ in equation 22, we find the operating characteristic of LRT as

$$P_{\mathrm{D}} = (2 - P_{\mathrm{F}})P_{\mathrm{F}} \tag{24}$$

(c) i. Because $P_{\mathrm{D}} = (2 - P_{\mathrm{F}})P_{\mathrm{F}}$, let $P_D \geqslant (1+\epsilon)P_{\mathrm{F}}$, we have $P_{\mathrm{F}} \leqslant (1-\epsilon)$. And P_{D} is a monotonically increasing function of P_{F} in [0,1], so $P_{\mathrm{D}}^{\max}(\epsilon) = 1 - \epsilon^2$

ii. $P_{\mathrm{D}}^{\max}(\epsilon) = 1 - \epsilon^2 > 0$ so $\epsilon \in (0, 1)$

iii. In the decision rule from part (a), $P_{\mathrm{F}} = 1 - \frac{p}{2(1-p)} \leqslant 1 - \epsilon$, that is $p \geqslant \frac{2\epsilon}{1+2\epsilon}$

3.5. (a) i. The cost is

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

So

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

Specifically,

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

α) $\hat{H} = H_1$: 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 $2\pi_3(\underline{y}) < 2\pi_1(\underline{y}) + \pi_2(\underline{y})$

β) $\hat{H} = H_2$: 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 $2\pi_3(\underline{y}) < \pi_1(\underline{y}) + 2\pi_2(\underline{y})$

γ) $\hat{H} = H_3$: 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,

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

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

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

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

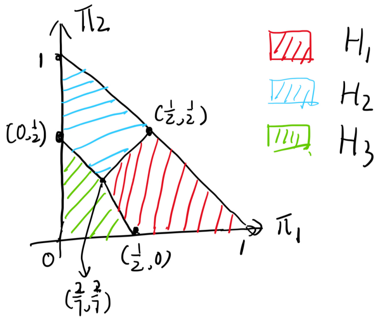

Figure 1: Decision regions in the (π_1, π_2) plane

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

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

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

$$L_i(\underline{y}) = \frac{p_{\underline{\mathrm{y}}|\mathrm{H}}\left(\underline{y} \mid H_i\right)}{p_{\underline{\mathrm{y}}|\mathrm{H}}\left(\underline{y} \mid H_1\right)} = \frac{\pi_i(\underline{y})}{\pi_1(\underline{y})} \tag{31}$$

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

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

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

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

So the optimum decision rule can be specfied in terms of the pair of suffcient statistics

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

And the optimum decision rule is,

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

3.6. (a) At the local sensor, $P(\mathrm{y_i} > 0|\mathrm{x} = 1) = P(\mathrm{y_i} < 0|\mathrm{x} = -1) = 0.75$, $P(\mathrm{y_i} < 0|\mathrm{x} = 1) = P(\mathrm{y_i} > 0|\mathrm{x} = -1) = 0.25$.
Thus, the conditional distribution of $\hat{x}_i$ given x is,

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

At the fusion center,

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

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

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

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

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

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

(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

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

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

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

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

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

which equals to,

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

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,

$$\begin{aligned} & \sum_{\hat{x}_2, x} p_{\hat{\mathrm{x}}_2,\mathrm{x}}(\hat{x}_2, x) p_{\mathrm{y}_1|\mathrm{x}}(y_1|x) \left[C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)\right] \\ = & \sum_{\hat{x}_2} p_{\hat{\mathrm{x}}_2,\mathrm{x}}(\hat{x}_2, 1) p_{\mathrm{y}_1|\mathrm{x}}(y_1|1) \left[C(1, \hat{x}_2, 1) - C(-1, \hat{x}_2, 1)\right] \\ & + \sum_{\hat{x}_2} p_{\hat{\mathrm{x}}_2,\mathrm{x}}(\hat{x}_2, -1) p_{\mathrm{y}_1|\mathrm{x}}(y_1|-1) \left[C(1, \hat{x}_2, -1) - C(-1, \hat{x}_2, -1)\right] \end{aligned} \tag{44}$$

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 $C(1, \hat{x}_2, -1) > C(-1, \hat{x}_2, -1)$.

$$\sum_{\hat{x}_2, x} p_{\hat{\mathrm{x}}_2,\mathrm{x}}(\hat{x}_2, x) p_{\mathrm{y}_1|\mathrm{x}}(y_1|x) \left[C(1, \hat{x}_2, x) - C(-1, \hat{x}_2, x)\right] < 0 \Longleftrightarrow \frac{p_{\mathrm{y}_1|\mathrm{x}}(y_1|1)}{p_{\mathrm{y}_1|\mathrm{x}}(y_1|-1)} > \gamma_1 \tag{45}$$

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

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

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

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

(c) In the same way,

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

where

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

(d) As the Equation 51 shows,

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

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

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

So when $L = 2, \gamma_1 = \frac{p_{\mathrm{x}}(x=-1)}{p_{\mathrm{x}}(x=1)}$ 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!