Lecture 4

Many Hypothesis Testing

1. Review of Binary Hypothesis Testing

Thus far we have studied the Bayesian Decision:

$$E_B(y) = \arg\min_i \sum_{j=0}^{M-1} C(i, j) \pi_j(y)$$

(4.1)

We remark that (4.1) is a very general result, of which we can derive two special cases:

(1) if \( C_{ii} = C_{oo} = 0 \) \; \( C_{io} = C_{oi} = 1 \)

(This is the uniform cost),

then the Bayesian decision rule becomes a likelihood ratio test (LRT)

$$L(y) \leq \frac{\pi_0}{\pi_1}$$

(4.2)

where

$$L(y) = \frac{f_1(y)}{f_0(y)}$$

(2) if, in addition to (1), we also have \( \pi_0 = \pi_1 \),

then (4.2) reduces to \( L(y) \leq 1 \). As long as the cost is uniform, Bayesian Decision rule is also a Maximum a Posteriori rule. In this case, the probability of error = Bayesian risk:

$$P(\theta \neq \delta(y)) = E_Y \left[ E_{\theta|Y}[C(\delta(y), \theta) | Y = y] \right]$$
The relationship between these rules is shown below:

General Bayesian Decision (Binary)

\[ \delta_B = \arg \min_{R_{Bk}} L(y) \]

\[ L(y) \leq \frac{H_0 \cdot \varepsilon \cdot \varepsilon}{H_1 \cdot \varepsilon \cdot \varepsilon} \frac{\pi_0}{\pi_1} \]

\[ \delta_B = \delta_{\text{MAP}} \]

\[ R_{Bk} = \text{Prob. Error} \]

Example: Binary Symmetric Channel

Consider a communication channel as follows:

\[ \begin{array}{c|c|c}
\pi_0 & 0 & 0 \\
\hline
\pi_1 & \lambda_0 & \lambda_1 \\
\end{array} \]

Suppose that the communication channel is symmetric, i.e. \( \lambda_0 = \lambda_1 = \lambda \), and assume a uniform cost, \( \pi_0 = \pi_1 \). Given an observation \( y \in \{0, 1\} \), find the Bayesian decision rule.

Solution (under the Bayesian decision framework, we know that \( \delta_B(y) = \) is to check

\[ L(y) \leq \frac{H_0}{H_1} \frac{\pi_0}{\pi_1} = 1 \]

Therefore, since

\[ L(y) = \frac{f_1(y)}{f_0(y)} = \frac{\frac{\pi_1}{\pi_0} \frac{1}{y} \cdot \lambda_1}{\frac{1}{y} \cdot \lambda_0} = \left\{ \begin{array}{ll}
\frac{\lambda_1}{\lambda_0}, & y = 0, \\
\frac{\lambda_0}{\lambda_1}, & y = 1
\end{array} \right. \]
We have:

if $y = 0$, \( \frac{\lambda}{1 - \lambda} \leq \frac{\lambda_0}{\lambda_1} \Rightarrow \lambda \leq \frac{\lambda_0}{\lambda_1} \geq \frac{1}{2} \)

So, \( \delta_B(0) = \begin{cases} 0 & \text{if } \lambda \leq \frac{1}{2} \\ 1 & \text{if } \lambda > \frac{1}{2} \end{cases} \)

If $y = 1$, \( \frac{1 - \lambda}{\lambda} \leq \frac{\lambda_0}{\lambda_1} \Rightarrow \lambda \geq \frac{\lambda_0}{\lambda_1} \leq \frac{1}{2} \)

So, \( \delta_B(1) = \begin{cases} 0 & \text{if } \lambda \geq \frac{1}{2} \\ 1 & \text{if } \lambda \leq \frac{1}{2} \end{cases} \)

So in general, we have

\[
\delta_B(y) = \begin{cases} y & \text{if } \lambda \leq \frac{1}{2} \\ 1 - y & \text{if } \lambda > \frac{1}{2} \end{cases}
\]  

(4.3)

Intuitively, what (4.3) implies is that (if \( \lambda \geq \frac{1}{2} \), then we need to "flip" the bit. Does it make sense physically?)

Finally, we can check the probability of error, which is

\[
P_e = \Pi_0 \Pr(\delta(y) \neq 0 \mid H_0) + \Pi_1 \Pr(\delta(y) \neq 1 \mid H_1)
\]

\[
= \begin{cases} \Pi_0 \lambda + \Pi_1 \lambda & \text{if } \lambda \leq \frac{1}{2} \\ \Pi_0 (1 - \lambda) + \Pi_1 (1 - \lambda) & \text{if } \lambda > \frac{1}{2} \end{cases}
\]

\[
P_e = \begin{cases} \lambda & \text{if } \lambda \leq \frac{1}{2} \\ 1 - \lambda & \text{if } \lambda > \frac{1}{2} \end{cases}
\]

If we do not flip the bit.
2. M-ary Hypothesis Testing

As a generalization of binary hypothesis testing, the M-ary hypothesis testing partitions the observation space $\mathcal{X}$ into $M$ partitions:

Mathematically, the M-ary decision rule is

$$
\delta_B(y) = \underset{i}{\text{argmin}} \sum_{j=0}^{M-1} C(i, j) \pi_j f_j(y)
$$

$$
= \underset{i}{\text{argmin}} \sum_{j=0}^{M-1} C(i, j) \frac{\pi_j f_j(y)}{\sum_{i=0}^{M-1} \pi_k f_k(y)}
$$

$$
= \underset{i}{\text{argmin}} \sum_{j=0}^{M-1} C(i, j) \pi_j \frac{f_j(y)}{f_0(y)}
$$

$$(\because \sum_{i=0}^{M-1} \pi_i f_i(y) \text{ is indep of } i)$$

$$(\because f_0(y) \text{ is indep of } i)$$

$$
\delta_B(y) = \underset{i}{\text{argmin}} \sum_{j=0}^{M-1} C(i, j) \pi_j L_j(y),
$$

where in (a) we define $L_j(y) = \frac{f_j(y)}{f_0(y)}$.

If we further define

$$
\hat{h}_i(y) = \sum_{j=0}^{M-1} C(i, j) \pi_j L_j(y),
$$

then (4.4) becomes

$$
\delta_B(y) = \underset{i}{\text{argmin}} \hat{h}_i(y),
$$

$$
(4.5)
$$
Pictorially, (4.5) suggests that \((M=3)\)

In the special case of uniform cost, the partition looks like follows:

\[
\begin{align*}
& \text{Cost:} \quad C_{00} = C_{11} = C_{22} = 0 \\
& C_{0i} = 1 \quad \text{if } i \neq j
\end{align*}
\]

Example:

Let \(\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\), and let \(N \sim \mathcal{N}(0, \sigma^2 I_3)\).

Suppose the hypothesis is:

\(H_i : \quad Y = \mu_i + N\)

Assume uniform cost, assume \(\Pi_0 = \frac{1}{4}, \Pi_1 = \frac{1}{4}, \Pi_2 = \frac{1}{2}\).

Find the Bayes decision rule.
Solution By (4.4) and (4.5), we know that

$$\delta_{\beta}(y) = \arg\min_{0 \leq j \leq 2} C_{00}\pi_0 + C_{01}\pi_{11}(y) + C_{02}\pi_2 L_2(y),$$

where we can also define

$$h_0(y) = C_{00}\pi_0 + C_{10}\pi_1 L_1(y) + C_{02}\pi_2 L_2(y),$$

$$h_1(y) = C_{10}\pi_0 + C_{11}\pi_1 L_1(y) + C_{12}\pi_2 L_2(y),$$

$$h_2(y) = C_{20}\pi_0 + C_{21}\pi_1 L_1(y) + C_{22}\pi_2 L_2(y).$$

With some calculations we can show that

$$h_0(y) = \frac{1}{4} L_1(y) + \frac{1}{2} L_2(y),$$

$$h_1(y) = \frac{1}{4} L_1(y) + \frac{1}{2} L_2(y),$$

$$h_2(y) = \frac{1}{4} L_1(y) + \frac{1}{2} L_1(y).$$

The M-ary hypothesis testing says we should claim

$$H_0$$ if $$h_0 \leq h_1$$ and $$h_0 \leq h_2$$. So we must have

$$h_0 \leq h_1 \Rightarrow \frac{1}{4} L_1(y) + \frac{1}{2} L_2(y) \leq \frac{1}{4} + \frac{1}{2} L_2(y) \Rightarrow \frac{1}{4} L_1(y) - \frac{1}{4} \leq 0 \Rightarrow L_1(y) \leq 1$$

$$h_0 \leq h_2 \Rightarrow \frac{1}{4} L_1(y) + \frac{1}{2} L_2(y) \leq \frac{1}{4} + \frac{1}{4} L_1(y) \Rightarrow \frac{1}{2} L_2(y) - \frac{1}{4} \leq 0 \Rightarrow L_2(y) \leq \frac{1}{2}$$

Now, let's compute $$L_1(y)$$ and $$L_2(y)$$. By definition, we know that

$$f_j(y) \overset{def}{=} f_{Y | H_j}(y) = \frac{1}{\sqrt{(2\pi)^n \sigma^2}} \exp\left(-\frac{||y - \mu_j||^2}{2 \sigma^2}\right).$$

(Remark: Multivariate Gaussian is

$$\frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2}\right),$$

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.9 \end{pmatrix}$$)
Therefore, the likelihood ratio is
\[ L_1(y) = \frac{f_l(y)}{f_0(y)} = \exp\left(\frac{-1}{2\sigma^2} \left( \| y - \mu \|_2^2 - \| y - \mu_0 \|_2^2 \right) \right) \]
\[ = \exp\left(\frac{-1}{2\sigma^2} \left( \| \begin{bmatrix} y_0^0 \\ y_0^1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \|_2^2 - \| \begin{bmatrix} y_0^0 \\ y_0^1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \|_2^2 \right) \right) \]
\[ = \exp\left(\frac{-1}{2\sigma^2} \left( (y_0^2 + (y_1 - 1)^2 + y_2^2) - ((y_0 - 1)^2 + y_1^2 + y_2^2) \right) \right) \]
\[ = \exp\left(\frac{-1}{2\sigma^2} \left( 2y_1 - 2y_1 \right) \right) = \exp\left(\frac{-1}{2\sigma^2} \right) \]
\[ = \exp\left(\frac{-y_1 - y_0}{G^2} \right) \]

Similarly,
\[ L_2(y) = \exp\left(\frac{y_2 - y_0}{G^2} \right) \]

Therefore, \( L_1(y) \leq 1 \) and \( L_2(y) \leq \frac{1}{2} \) implies
\[ \exp\left(\frac{y_1 - y_0}{G^2} \right) \leq 1 \quad \text{and} \quad \exp\left(\frac{y_2 - y_0}{G^2} \right) \leq \frac{1}{2} \]
\[ \Rightarrow \begin{cases} y_1 \leq y_0 \\ y_2 \leq y_0 - 2\sigma^2 \ln 2 \end{cases} \]

And you can do the rest, following the same argument.

Remark: For any vector \( y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix} \), the Euclidean norm (or the \( L_2 \)-norm) is defined as
\[ \| y \|_2^2 = y_0^2 + y_1^2 + y_2^2 \]
In general \( \| y \|_2^2 = \sum_{j=0}^{M-1} y_j^2 \) for any \( M \)-vector.