Lecture 3
Sept. 22, 2010

Composite Hypothesis Testing

In previous lectures, we considered what may be termed Simple Hypothesis Testing where for each hypothesis, we assume that one distribution is at work. For example under H_0: a Gaussian with mean μ_0 and under H_1: a Gaussian of mean μ_1. Sometimes, we need to consi[der] many possible distributions. For example

$$H_0 : y \sim N(0, \sigma^2)$$

versus

$$H_1 : y \sim N(m, \sigma^2)$$

where $m_1 \leq m \leq m_2$.

This type of hypothesis testing is called Composite Hypothesis Test.

To model <u>Composite</u> Hypothesis test, we consider a family of distributions indexed by a parameter $\theta \in \Lambda$ where Λ is the parameter set. For the above example $\theta = m$ and $\Lambda = [m_1, m_2]$. Denote this set of distributions as $\{P_\theta : \theta \in \Lambda\}$ or equivalently (for densities):

$$\{p(y|\theta) : \theta \in \Lambda\}$$

To complete the model we need a <u>cost</u> function of the form $C[i, \theta]$ where

$C[i, \theta]$ is the cost of deciding $i = 0, 1$ when $Y \sim P_\theta$, $\theta \in \Lambda$.

For example in radar problem, the position is found from the time measurement and velocity is calculated based on Frequency (Doppler shift) and θ can represent position or velocity and H_0, H_1 may be defined as fast versus slow or close versus far (or absent versus present).

Consider a decision rule δ. Then

$$R_\theta(\delta) = E_\theta\{C[\delta(Y),\theta]\} \quad \theta \in \Lambda$$

$E_\theta \triangleq$ expectation assuming $Y \sim P_\theta$.

The Bayes (average) risk) is:

$$r(\delta) = E\{R_\theta(\delta)\}$$
$$= E[E\{C[(Y),\theta]|\theta\}]$$
$$= E\{C[\delta(Y),\theta]\}$$

That is, $r(\delta)$ is cost of using δ averaged over Y and θ.

We can write $r(\delta)$ also as

$$r(\delta) = E\{E\{C[\delta(Y),\theta]|Y\}\}$$

From this, we see that $r(\delta)$ is minimiz if for

each $y \in \Gamma$ we choose $\delta(y)$ that minimizes

$$E\{C[\delta(y),\theta]|Y=y\}$$

Since $\delta(Y)$ can only take 0 or 1,

$$\delta_B(y) = \begin{cases} 1 & \text{if } E\{C[1,\Theta]|Y=y\} < E\{C[0,\Theta]|Y=y\} \\ 0 \text{ or } 1 & \text{if } E\{C[1,\Theta]|Y=y\} = E\{C[0,\Theta]|Y=y\} \\ 0 & \text{if } E\{C[1,\Theta]|Y=y\} > E\{C[0,\Theta]|Y=y\}. \end{cases}$$

When the whole parameter space can be segmented into two disjoint parts Λ_0 and Λ_1 one representing H_0 and another H_1, and with uniform cost, i.e.,

$$C[i, \theta] = C_{ij} \qquad \theta \in \Lambda_j$$

Then:

$$\delta_B(y) = \begin{cases} 1 & \\ 0 \text{ or } 1 & \text{if } \dfrac{P(\Theta \in \Lambda_1|Y=y)}{P(\Theta \in \Lambda_0|Y=y)} \begin{matrix} > \\ = \\ < \end{matrix} \dfrac{C_{10}-C_{00}}{C_{01}-C_{11}}, \\ 0 & \end{cases}$$

where $P(\theta \in \Lambda_j | Y=y)$ is probability that θ lies in Λ_j conditioned on $Y=y$.

$$P(\theta \in \Lambda_j | Y=y) = \frac{p(y|\theta \in \Lambda_j)P(\theta \in \Lambda_j)}{p(y)}$$

Thus:

$$\delta_B(y) = \begin{cases} 1 & \\ 0 \text{ or } 1 & \text{if } L(y) \begin{matrix} > \\ = \\ < \end{matrix} \dfrac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})} \\ 0 & \end{cases}$$

where, the likelihood ration $L(y)$ is

$$L(y) = \frac{p(y \mid \theta \in \Delta_1)}{p(y \mid \theta \in \Delta_0)}$$

Further, we can write

$$p(y \mid \theta \in \Delta_j) = \int_{\Delta_j} p(y \mid \theta)\, p(\theta \mid \theta \in \Delta_j)\, d\theta$$

So,

$$L(y) = \frac{\int_{\Delta_1} p(y \mid \theta)\, p(\theta \mid \Delta_1)\, d\theta}{\int_{\Delta_0} p(y \mid \theta)\, p(\theta \mid \Delta_0)\, d\theta}$$

Signal Detection:

$$H_0 : \; Y_k = N_k + S_{0k} \qquad k=1,2,\dots,n$$
$$H_1 : \; Y_k = N_k + S_{1k} \qquad k=1,2,\dots,n$$

Let

$\underline{Y} = (Y_1, \dots, Y_n)$ observation vector

$\underline{N} = (N_1, \dots, N_n)$ noise vector

and

$S_0 = (S_{01}, \dots S_{0n}) \rightarrow H_0$

and

$S_1 = (S_{11}, \dots S_{1n}) \rightarrow H_1$

are vectors of samples from two signals.

assume

$$\underline{N} \sim P_N(\underline{n}) \Rightarrow p(\underline{y} \mid \underline{s}_i) = P_N(\underline{y} - \underline{s}_i)$$

then

$$p(\underline{y} \mid H_i) = E\{P_N(\underline{y} - \underline{s}_i)\}$$

So,

$$L(\underline{y}) = \frac{E\{P_N(\underline{y} - \underline{s}_1)\}}{E\{P_N(\underline{y} - \underline{s}_0)\}}$$

Detection of deterministic Signals:

In most Communication cases the signal vectors $\underline{S}_0$ and $\underline{S}_1$ are known (deterministic). So,

$$L(\underline{y}) = \frac{p_{\underline{N}}(\underline{y} - \underline{S}_1)}{p_{\underline{N}}(\underline{y} - \underline{S}_0)}$$

$$= \frac{p_{\underline{N}}(y_1 - s_{11}, \cdots, y_n - s_{1n})}{p_{\underline{N}}(y_1 - s_{01}, \cdots, y_n - s_{0n})}$$

When the noise samples are independent,

$$p_{\underline{N}}(\underline{y}) = \prod_{k=1}^{n} p_{N_k}(y_k)$$

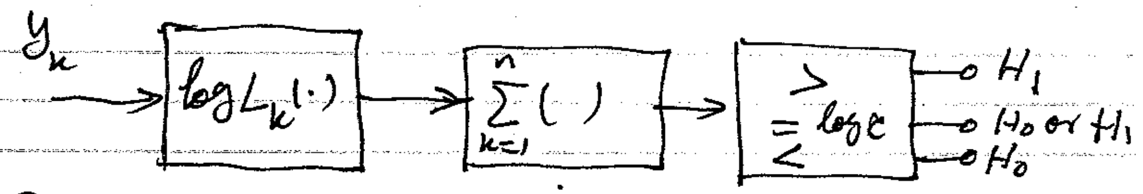
and

$$L(\underline{y}) = \prod_{k=1}^{n} L_k(y_k)$$

where $$L_k(y_k) = \frac{p_{N_k}(y_k - s_{1k})}{p_{N_k}(y_k - s_{0k})}$$

Thus, the optimum tests will be:

$$\tilde{\delta}_0(y) = \begin{cases} 1 & \text{if } \sum_{k=1}^{n} \log L_k(y_k) > \log \tau \\ \gamma & \text{if } \sum_{k=1}^{n} \log L_k(y_k) = \log \tau \\ 0 & \text{if } \sum_{k=1}^{n} \log L_k(y_k) < \log \tau \end{cases}$$



Example: ~~Detection~~ in i.i.d. Gaussian noise

$$N_k \sim N(0, \sigma^2) \quad \forall k.$$

Let $\underline{s}_0 = (0, 0, \cdots, 0)$

and

$$\underline{s}_1 = (s_1, \cdots, s_n)$$

Note: any $\underline{s}_0 = (s_{01}, \cdots s_{0n})$

$$\underline{s}_1 = (s_{11}, \cdots s_{1n})$$

case can be dealt with by letting.

$$\underline{s}_0' = (0, 0, \cdots, 0)$$

$$\underline{s}_1' = (s_{11} - s_{01}, s_{12} - s_{02}, \cdots, s_{1n} - s_{0n})$$

$$= (s_1, s_2, \cdots s_n)$$

$$p(y_k | \underline{s}_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y_k^2/2\sigma^2}$$

$$p(y_k | \underline{s}_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y_k - s_k)^2/2\sigma^2}$$

$$L_k(y_n) = e^{\frac{s_k(y_k - s_k/2)}{\sigma^2}}$$

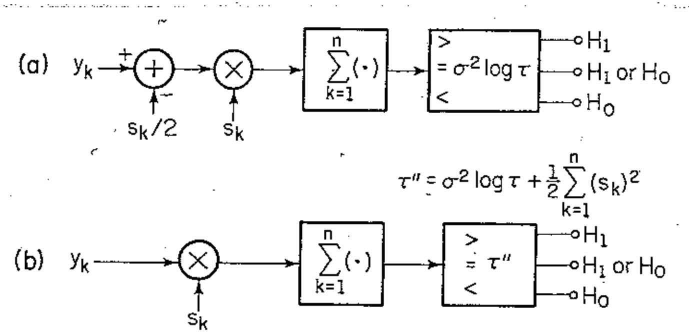
or

$$\log L_k(y_k) = \frac{s_k(y_k - s_k/2)}{\sigma^2}$$

and the decision rule is:

$$\tilde{\delta}_0(y) = \begin{cases} 1 & \\ \gamma & \sum_{k=1}^{n} s_k(y_k - s_k/2) \quad \begin{matrix} > \tau' \\ = \tau' \\ < \tau' \end{matrix} \\ 0 & \end{cases}$$

where $\tau' = \sigma^2 \log \tau$.



Optimum detector for coherent signals i.i.d. Gaussian noise.