Bayes decision rule (continued)

In the last section we noted that the Bayes decision rule can be written as:

$$I_k(R) \geq_{not \ H_m} I_m(R)$$

which means that our task is to select the hypothesis with the minimum $$I_i(R)$$.

The above decision rule is re-written as:

$$\sum_j P_j(C_{kj} - C_{jj})p(R|H_j) \geq_{not \ H_m} \sum_j P_j(C_{mj} - C_{jj})p(R|H_j)$$

After cancelling the common terms, we obtain

$$(C_{km} - C_{mm})P_m p(R|H_m) \geq_{not \ H_m} (C_{mk} - C_{kk})P_k p(R|H_k) + \sum_{j \neq m, j \neq k} (C_{mj} - C_{kj})P_j p(R|H_j)$$

Define the $$i$$th likelihood function via:

$$\Lambda_i(R) \triangleq \frac{p(R|H_i)}{p(R|H_0)}$$

Therefore, the test becomes:

$$(C_{km} - C_{mm})P_m \Lambda_m(R) \geq_{not \ H_m} (C_{mk} - C_{kk})P_k \Lambda_k(R) + \sum_{j \neq m, j \neq k} (C_{mj} - C_{kj})P_j \Lambda_j(R)$$

- Note that there are $$\frac{M(M-1)}{2}$$ such inequalities.
- Note that the decision space is the $$(M - 1)$$ dimensional space of the likelihood function.
Figure 2. Decision regions

Special cases - Minimum probability of error criterion:

Cost functions are defined via:

$$C_{ij} = \begin{cases} 1 & ; i \neq j \\ 0 & ; i = j \end{cases}$$

i.e. unit cost function for all wrong decisions.

In this case, the risk function becomes:

$$R = \sum_j P_j \sum_i C_{ij} \int_{Z_i} p(\bar{R}|H_j)d\bar{R}$$

$$= \sum_j P_j \sum_{i \neq j} \int_{Z_i} p(\bar{R}|H_j)d\bar{R}$$

$$\underbrace{\text{Prob}[\text{choose } H_i | H_j \text{ is true}]}_{\text{Prob}[\text{error } | H_j \text{ is true}]} = \sum_j \text{Prob}[\text{error } | H_j \text{ is true}] = P(\text{error})$$

- Thus, minimizing the risk function with $C_{ij} = \delta_{ij}$ is equivalent to minimizing the overall probability (expected value) of error percentage.

After substituting $C_{ij} = \delta_{ij}$ in

$$I_k(\bar{R}) \preceq_{\text{not } H_m} I_m(\bar{R})$$

we obtain:

$$P_m \Lambda_m(\bar{R}) \preceq_{\text{not } H_m} P_k \Lambda_k(\bar{R})$$

$$P_m p(\bar{R}|H_m) \preceq_{\text{not } H_m} P_k p(\bar{R}|H_k)$$

Thus, for a given measurement $\bar{R}$, select the hypothesis that maximizes

$$P_i \Lambda_i(\bar{R}), \forall i = 0, 1, \ldots, M - 1$$
Rewrite equation (1) using the log-likelihood function:
\[ \ln P_m + \ln \Lambda_m(\hat{R}) \overset{\text{mut}}{\gtrless} \ln P_k + \ln \Lambda_k(\hat{R}) \]

Going back to the original channel pdf's, we may also write the decision rule via:
\[ P_m p(\hat{R}|H_m) \overset{\text{mut}}{\gtrless} P_k p(\hat{R}|H_k) \]

Dividing both sides by \( p(\hat{R}) \), where \( p(\hat{R}) = \sum_{j=0}^{M-1} p(\hat{R}|H_j) \), we get:
\[ \frac{P_m p(\hat{R}|H_m)}{p(\hat{R})} \overset{\text{mut}}{\gtrless} \frac{P_k p(\hat{R}|H_k)}{p(\hat{R})} \]

Recall: \( p(a|b) = \frac{p(a)p(b|a)}{p(b)} \)
Associate \( a \to H_m, b \to \hat{R} \)
\[ \therefore p(H_m|\hat{R}) \overset{\text{mut}}{\gtrless} \frac{p(H_k|\hat{R})}{p(\hat{R})} \]

\[ \therefore \text{Therefore, the decision rule is to select the hypothesis that yields the maximum a posteriori pdf (i.e. the probability of a hypothesis given an observation).} \]

This is called the Maximum A Posteriori Probability (MAP) decision rule.

Summary
Minimum probability of error decision rule and MAP decision rule are the same; they are special cases of the Bayes decision rule with uniform cost for all incorrect decisions.

Example
Binary decision, i.e \( M = 2 \) ⇒ only one likelihood function.
\[ \Lambda(\hat{R}) = \frac{p(\hat{R}|H_1)}{p(\hat{R}|H_0)} \triangleq \Lambda(\hat{R}) \]
\[ \therefore \text{The Bayes decision becomes} \]
\[ \Lambda(\hat{R}) \overset{\text{select } H_1}{\gtrless} \frac{p_0(C_{10} - C_{00})}{p_1(C_{01} - C_{11})} \triangleq \eta \]
\( \eta \), the threshold is set by the user based on the a priori probabilities (i.e \( P_0 \) and \( P_1 = 1 - P_0 \)) and the assigned cost functions.

Receiver
The receiver structure is as shown in figure 3.

Special case
MAP decision rule: substitute \( C_{00} = C_{10} = 0 \) and \( C_{11} = C_{01} = 1 \).
\[ \Rightarrow \eta = \frac{P_0(1 - 0)}{P_1(1 - 0)} = \frac{P_0}{P_1} = \frac{P_0}{(1 - P_0)} \]
Log likelihood form:
\[ L(\hat{R}) \triangleq \ln \Lambda(\hat{R}) \overset{\text{select } H_1}{\gtrless} \ln \eta \triangleq \xi \]
where

$$\xi = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

e.g.

$$H_0 : \vec{R} = \vec{S}_0 + \vec{N}$$
$$H_1 : \vec{R} = \vec{S}_1 + \vec{N}$$

where

$$\vec{S}_0 = \begin{bmatrix} S_{01} \\ S_{02} \\ \vdots \\ S_{0N} \end{bmatrix}, \vec{S}_1 = \begin{bmatrix} S_{11} \\ S_{12} \\ \vdots \\ S_{1N} \end{bmatrix}, \vec{N} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{bmatrix}$$

Source and channel model

$n_i$’s are independent identically distributed (iid) normal or Gaussian r.v.’s with zero mean and variance $\sigma^2_N$; $n_i \sim N(0, \sigma^2_N)$, $\forall i$.

$$E(n_i n_j) = \sigma^2_N \delta_{ij}$$
$$E(n_i) = 0$$

$$p(\vec{N}) = \frac{1}{(\sqrt{2\pi} \sigma_N)^N} \exp \left[ - \frac{\sum_{i=1}^N n_i^2}{2\sigma^2_N} \right]$$

$S_0$ and $S_1$ are deterministic (known) vectors, e.g. samples of two known signals.

The observation pdf:

$$H_k : \vec{R} = \vec{S}_k + \vec{N}; k = 0, 1$$

$\Rightarrow \vec{R}|H_k$ is also normal multivariate with mean $E(\vec{R}|H_k) = \vec{S}_k$ and covariance matrix:

$$Cov = \begin{bmatrix} \sigma_n^2 & 0 & \cdots & 0 \\ 0 & \sigma_n^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$
In this case its pdf is:

\[ p(\vec{R} | H_k) = \frac{1}{(\sqrt{2\pi}\sigma_N)^N} \exp \left[ -\frac{\sum_{i=1}^N (r_i - S_{ki})^2}{2\sigma_n^2} \right] \]

where

\[ \vec{R} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} \]

Likelihood function:

\[ \Lambda(\vec{R}) = \frac{p(\vec{R} | H_1)}{p(\vec{R} | H_0)} = \frac{\exp \left[ -\sum_{i=1}^N (r_i - S_{1i})^2/2\sigma_n^2 \right]}{\exp \left[ -\sum_{i=1}^N (r_i - S_{0i})^2/2\sigma_n^2 \right]} \]

Log-likelihood function:

\[ L(\vec{R}) = \ln(\Lambda(\vec{R})) \Rightarrow \sum_{i=1}^N 2r_i(S_{1i} - S_{0i})/2\sigma_n^2 \gtrless \xi \sigma_n^2 + 1/2 \sum_{i=1}^N S_{1i}^2 - 1/2 \sum_{i=1}^N S_{0i}^2 \]

Rewrite the above via:

\[ \sum_{i=1}^N 2r_i(S_{1i} - S_{0i})/2\sigma_n^2 \gtrless \xi \sigma_n^2 + 1/2 \sum_{i=1}^N S_{1i}^2 - 1/2 \sum_{i=1}^N S_{0i}^2 \]

Define

\[ E_0 \triangleq \sum_{i=1}^N S_{0i}^2 \]: energy of \( \vec{S}_0 \)

\[ E_1 \triangleq \sum_{i=1}^N S_{1i}^2 \]: energy of \( \vec{S}_1 \)

and note that \( \sum_{i=1}^N r_i S_{ki} = \vec{R}^T \vec{S}_k \triangleq < \vec{R}, \vec{S}_k > \), also known as the projection of \( S_k \) onto \( \vec{R} \). The decision equation becomes:

\[ (S_1 - S_0)^T \vec{R} \gtrless \xi \sigma_n^2 + 1/2(E_1 - E_0) \]

Note that \( E_k = < \vec{S}_k, \vec{S}_k > = \vec{S}_k^T \vec{S}_k \).

Therefore the projection of \( \vec{R} \) into \( \vec{S}_1 - \vec{S}_0 \) is the only information required (sufficient statistic) for decision making.

**Unipolar ASK:**

\[ \vec{S}_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \vec{S}_1 = \begin{bmatrix} A \\ A \\ \vdots \\ A \end{bmatrix} \]

To transmit 0 or 1, an ON/OFF signaling scheme is required.

Therefore, the decision rule becomes:

\[ \sum_{i=1}^N r_i(A - 0) \gtrless \xi \sigma_n^2 + 1/2(NA^2 - 0) \]
Dividing both sides by $N.A$

$$\frac{1}{N} \sum_{i=1}^{N} r_i \leftrightarrow \begin{cases} H_1 : \bar{R} \sim N(A, \frac{\sigma^2}{N}), r_i = A + n_i \\ H_0 : \bar{R} \sim N(0, \frac{\sigma^2}{N}), r_i = 0 + n_i \end{cases}$$

\[ \bar{R} \leftrightarrow \begin{cases} \text{select } H_1 \leftrightarrow \frac{\xi \sigma^2}{\sqrt{N}} + \frac{A}{2} \triangleq \gamma \\ \text{select } H_0 \rightarrow \end{cases} \]

Note that for the decision, it is sufficient to reduce the processing by using the scalar $\bar{R}$ instead of the $N$-dimensional $\vec{R}$.

Pdf of $\bar{R}$ is given by:

$$H_0 : \bar{R} \sim N(0, \frac{\sigma^2}{N}), r_i = 0 + n_i$$

$$H_1 : \bar{R} \sim N(A, \frac{\sigma^2}{N}), r_i = A + n_i$$

$$\Rightarrow p(\bar{R}|H_0) = \frac{1}{\sqrt{2\pi} \frac{\sigma^2}{\sqrt{N}}} e^{\exp \left[ - \frac{\bar{R}^2}{2 \frac{\sigma^2}{N}} \right]}$$
\[ p(\bar{R}|H_1) = \frac{1}{\sqrt{2\pi \frac{\sigma_n^2}{N}}} e^{\exp \left[ - \frac{\bar{R} - A)^2}{2 \frac{\sigma_n^2}{N}} \right]} \]

Special case: MAP or min. probability of error criterion with \( P_0 = P_1 = \frac{1}{2} \) yields:

\[ \eta = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} = 1 \Rightarrow \xi = 0 \]

Probability of false alarm:

\[ P_F = P(\text{error}|H_0) = \int_{-\gamma}^{\gamma} p(\bar{R}|H_0)d\bar{R} = \text{erfc}^{*}\left(\frac{\gamma}{\sqrt{N}}\right) \]

Probability of miss:

\[ P_M = P(\text{error}|H_1) = \int_{-\gamma}^{\gamma} p(\bar{R}|H_1)d\bar{R} = \text{erfc}^{*}\left(\frac{A + r}{\sqrt{N}}\right) \]

Probability of detection:

\[ P_D = 1 - P_M = \int_{\gamma}^{\infty} p(\bar{R}|H_1)d\bar{R} \]

Probability of error:

\[ P_E = P_0.P(\text{error}|H_0) + P_1.P(\text{error}|H_1) \]

Bipolar ASK:

\[ \bar{S}_0 = \begin{bmatrix} -A \\ -A \\ \vdots \\ -A \end{bmatrix}, \bar{S}_1 = \begin{bmatrix} A \\ A \\ \vdots \\ A \end{bmatrix} \]

\[ \Rightarrow E_0 = NA^2, E_1 = NA^2. \text{ Therefore, the decision equation becomes:} \]

\[ \sum_{i=1}^{N} r_i[A - (-A)] \gtrless_{\text{select } H_1}^{\text{select } H_0} \xi \sigma_n^2 + \frac{1}{2}(NA^2 - NA^2) \]

or

\[ 2A \sum_{i=1}^{N} r_i \gtrless_{\text{select } H_1}^{\text{select } H_0} \xi \sigma_n^2 \]

or

\[ \bar{R} \gtrless_{\text{select } H_1}^{\text{select } H_0} \frac{\xi \sigma_n^2}{2AN} \triangleq \gamma \]

Pdf of \( \bar{R} \) is given by:

\[ H_0 : \bar{R} \sim N(-A, \frac{\sigma_n^2}{N}), r_i = 0 + n_i \]

\[ H_1 : \bar{R} \sim N(A, \frac{\sigma_n^2}{N}), r_i = A + n_i \]
\[ p(\vec{R}|H_k) = \frac{1}{\sqrt{2\pi} \sigma_k} \exp\left[ -\frac{(\vec{R} - S)^2}{2\sigma_k^2} \right] \]

where

\[ S = \begin{cases} -A & \text{under } H_0 \\ A & \text{under } H_1 \end{cases} \]

For \( \gamma = 0 \) or MAP or min probability of error with \( C_{ij} = \delta_{ij} \),

\[ P(\text{error}|H_0) = P(\text{error}|H_1) : \text{Bipolar} < P(\text{error}|H_0) = P(\text{error}|H_1) : \text{Unipolar} \]

Average transmitted energy:

\[ E_b \triangleq P_0 E_0 + P_1 E_1 \]

For \( P_0 = P_1 = \frac{1}{2} \)

\[ E_b = \begin{cases} \frac{1}{2} 0 + \frac{1}{2} N A^2 = \frac{NA^2}{2} & : \text{unipolar} \\ \frac{1}{2} \frac{NA^2}{2} + \frac{1}{2} \frac{NA^2}{2} = \frac{NA^2}{2} & : \text{bipolar} \end{cases} \]

Bipolar is better in probability of error performance since we use more energy in the transmission. Even if we adjust the \( A \) value to use the same average energy in both cases, still the bipolar performs better. The only scenario where the unipolar scheme is preferable are the asynchronous (non-coherent) systems.

**Sufficient statistic**

- In the previous section, we formulated testing of hypotheses based on an \( N \)-dimensional observed vector \( \vec{R} \).
- In an example, we demonstrated how the average value of the elements of \( \vec{R} \) is sufficient for the receiver to make a decision.
- We now present the general concept for what is referred to as sufficient statistic in decision theory.

Consider the transformation of \( \vec{R} \) denoted by:

\[ \vec{W}_{L \times 1} = T[\vec{R}_{N \times 1}] \]

where \( L \leq N \); thus the transformation is not necessarily reversible. For the time being assume \( W \) to be of dimension \( N \times 1 \) and that the inverse of \( T[\cdot] \) exists.

We partition \( \vec{W} \) into two parts:

\[ \vec{W}_{N \times 1} = [\vec{W}_{1L \times 1}, \vec{W}_{2(N-L) \times 1}] \]

We can write the likelihood function via:

\[ \Lambda \vec{R} \triangleq \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)} \]

We also know that if \( y = g(x) \), then

\[ p_Y(y) = \frac{p_X(x)}{\frac{dg(x)}{dx}} \]

If \( X \to \vec{W} \) and \( Y \to \vec{R} \), then

\[ p(\vec{R}|H_i) = \frac{p(\vec{W}|H_i)}{J} \]

where \( J \) is the Jacobian of the transformation from \( \vec{W} \) to \( \vec{R} \). The Jacobian function is invariant of \( H_i \). Substitute in the likelihood function in terms of functions of \( \vec{W} \):

\[ \Lambda \vec{R} = \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)} \]
Thus the test can be performed via processing the likelihood function for \( \mathbf{W} \). Using the Bayes theorem, we have:

\[
\begin{align*}
\mathcal{L}_{\theta}(\mathbf{R}) &= \frac{\mathcal{L}(\mathbf{W})}{\mathcal{L}(\mathbf{W})} \\
&= \frac{p(\mathbf{W}|H_1)}{p(\mathbf{W}|H_0)}
\end{align*}
\]

Thus the test can be performed via processing the likelihood function for \( \mathbf{W} \). Using the Bayes theorem, we have:

\[
\begin{align*}
p(\mathbf{W}|H_i) &= p(\mathbf{W}_1, \mathbf{W}_2|H_i) \\
&= p(\mathbf{W}_1|H_i) p(\mathbf{W}_2|\mathbf{W}_1, H_i)
\end{align*}
\]

Suppose there exists a partitioning of \( \mathbf{W} \) such that

\[
p(\mathbf{W}_2|\mathbf{W}_2, H_i) = p(\mathbf{W}_2|\mathbf{W}_1)
\]

i.e. it is invariant in the hypotheses. Using this in the likelihood function

\[
\begin{align*}
\mathcal{L}_{\theta}(\mathbf{R}) &= \frac{\mathcal{L}(\mathbf{W})}{\mathcal{L}(\mathbf{W})} \\
&= \frac{p(\mathbf{W}_1|H_1) p(\mathbf{W}_2|\mathbf{W}_1, H_1)}{p(\mathbf{W}_1|H_0) p(\mathbf{W}_2|\mathbf{W}_2, H_0)} \\
&= \frac{p(\mathbf{W}_1|H_1)}{p(\mathbf{W}_1|H_0)} \triangleq \mathcal{L}_{\theta}(\mathbf{W}_1) \geq \text{select } H_1 \text{ or } H_0 \eta
\end{align*}
\]

i.e the test is invariant in \( \mathbf{W}_2 \). In this case \( \mathbf{W}_1 \) is called sufficient statistic to construct the test.

e.g.: Binary hypothesis testing

\[
\begin{align*}
H_0 : r_i &= n_i \\
H_1 : r_i &= n_i + A
\end{align*}
\]

where \( i = 1, 2, \cdots, N \) and \( n_i \sim N(0, \sigma^2_n) \) and i.i.d \( \forall i \). \( A \) is a constant. The sufficient statistic is \( \mathbf{w}_1 = \sum_{i=1}^{N} r_i \).

It does not matter what \( \mathbf{W}_2 \) is.

\[
\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix} = \begin{bmatrix}
\frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & & & \ddots & \\
0 & \cdots & 0 & \cdots & 1
\end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_N \end{bmatrix}
\]

e.g.

\[
\begin{align*}
H_0 : \bar{R} &= S_0 + \bar{N} \\
H_1 : \bar{R} &= S_1 + \bar{N}
\end{align*}
\]

where \( \bar{N} \sim N(0, \sigma^2_n) \). \( S_0 \) and \( S_1 \) are constants. In that case, we showed that the decision is based on

\[
l(\mathbf{R}) \triangleq (S_1 - S_0)^T \mathbf{R} \geq \text{select } H_1 \text{ or } H_0 \text{ threshold}
\]

The threshold is a scalar that is sufficient statistic for this detection problem.