Part III. Decision and Estimation Theories

§§ Decision Theory: Fundamentals

* A decision theory problem can be conceptually formulated as
(for a fixed time period)

- The source generates an output out of M choices, referred to as hypotheses labeled $H_0, H_1, \ldots, H_{M-1}$.

  <ex> In a binary digital communication system (M=2),
  
  $H_0 \rightarrow$ the hypothesis that “zero” is sent.
  $H_1 \rightarrow$ the hypothesis that “one” is sent.

  <ex> In a radar or sonar system (M=2),
  
  $H_0 \rightarrow$ the hypothesis that the target is absent.
  $H_1 \rightarrow$ the hypothesis that the target is present.

  <ex> In a M-ary modulation system,
  
  $H_i \rightarrow$ the hypothesis that the $i$th symbol waveform is sent.

- The probabilistic transition mechanism (PTM) introduces randomness into the transmission. It is commonly known as “noisy channel” in digital communications.

  <ex> Additive Noise Channel (e.g., AWGN, random interference, random jamming)

  <ex> Multiplicative Noise Channel (e.g., nonselective fading)

\[ <ex> \text{Additive Noise Channel (e.g., AWGN, random interference, random jamming)} \]

\[ <ex> \text{Multiplicative Noise Channel (e.g., nonselective fading)} \]
Random Filtering Channel (e.g., dispersive fading)

- \( \mathbf{r}(\mu) \) is the received N-dimensional vector and \( f(\mathbf{r}|H_i) \) is the conditional probability density of \( \mathbf{r}(\mu) \) given that \( H_i \) is sent.

- The observation space \( \mathbb{D} \) is the domain viewed by the detector (or, communication receiver).

The decision theory provides with the detector a theoretical framework from which the observation space is partitioned into \( M \) regions, \( D_0, D_1, \ldots, D_{M-1} \). This forms the “decision rule.” If the received \( \mathbf{r}(\mu) \) falls in \( D_i \), the detector decides that \( H_i \) is true.

In a M-ary modulation system, the demodulator implements the decision rule that determines which symbol waveform is currently transmitted.
§ Binary Decision (M=2)

* Consider

\[ y_r(\mu) = [r_1(\mu), r_2(\mu), \ldots, r_N(\mu)] \]

is the received random vector.

PTM can be characterized by \( f(r|H_i), i=0,1 \), which is the conditional probability density of \( r(\mu) \), given that hypothesis \( H_i \) is sent.

\( D_i, i=0,1, \) are N-dim regions, \( D_0 \cup D_1 = \mathbb{D} \).

Given \( r(\mu) = r \), the decision rule is to determine

\[ H_0 \text{ if } r \in D_0 \]
\[ H_1 \text{ if } r \in D_1 \]

* The aim of the decision theory is to determine \( D_0 \) and \( D_1 \) such that a certain “optimality” can be achieved. That is, the choice of \( D_0 \) and \( D_1 \) has to optimize a “pre-determined” performance measure (pm). Such a process is called a criterion.

* Terminology:

\( P(H_i) \) is called an *a priori* probability.

\( P(H_i) \equiv \text{prob. that } H_i \text{ is sent.} \)

\( P(H_i|r) \) is called an *a posteriori* probability.

\( P(H_i|r) \equiv \text{prob. that } H_i \text{ is sent given that } r(\mu) = r \text{ is received.} \)

\( f(r|H_i) \) is called a likelihood density.

\( f(r|H_i) \equiv \text{prob. density of } r(\mu) \text{ given that } H_i \text{ is sent.} \)

* Bayes Criterion:

* Define the cost

\( C_{ij} \equiv \text{the cost that “} H_j \text{ is sent” and “} H_i \text{ is determined”} \)

Note: \( C_{ij} \geq 0 \) by default
• The Bayes criterion says that, given \( \{C_{ij}\} \), the decision rule should minimize the cost on the average.

Let \( \bar{C} \equiv \text{average cost} \)

\[
\bar{C} = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} P_{r}\left\{ \text{send } H_{j} \text{ "and" decide } H_{i}\right\} \\
= \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} P_{r}\left\{ \text{send } H_{j}\right\} P_{r}\left\{ \text{decide } H_{i} \text{ } | \text{ send } H_{j}\right\} \\
= C_{00} P(H_{0}) \int_{D_{0}} f(r \mid H_{0}) \text{d}r + C_{01} P(H_{0}) \int_{D_{0}} f(r \mid H_{1}) \text{d}r \\
+ C_{10} P(H_{0}) \int_{D_{1}} f(r \mid H_{0}) \text{d}r + C_{11} P(H_{1}) \int_{D_{1}} f(r \mid H_{1}) \text{d}r \\
\]

Assumptions: \( C_{10} > C_{00} \) and \( C_{01} > C_{11} \)

(The cost of a wrong decision is higher than that of a correct one.)

\[
\bar{C} = C_{00} P(H_{0}) \int_{D_{0}} f_{D_{0}}(r \mid H_{0}) \text{d}r + C_{01} P(H_{0}) \int_{D_{0}} f_{D_{0}}(r \mid H_{1}) \text{d}r \\
+ C_{10} P(H_{0}) \int_{D_{0}} f_{D_{0}}(r \mid H_{0}) \text{d}r + C_{11} P(H_{1}) \int_{D_{1}} f_{D_{1}}(r \mid H_{1}) \text{d}r \\
= P(H_{0})(C_{10} + P(H_{1})C_{11} + P(H_{0})(C_{00} - C_{10})) \int_{D_{0}} f_{D_{0}}(r \mid H_{0}) \text{d}r \\
+ P(H_{1})(C_{01} - C_{11}) \int_{D_{0}} f_{D_{0}}(r \mid H_{1}) \text{d}r \\
= P(H_{0})C_{10} + P(H_{1})C_{11} \\
+ P(H_{1})(C_{01} - C_{11}) \int_{D_{0}} f_{D_{0}}(r \mid H_{1}) \text{d}r - P(H_{0})(C_{10} - C_{00}) \int_{D_{0}} f_{D_{0}}(r \mid H_{0}) \text{d}r \\
\]

\( \bar{C} \) can be minimized if we choose \( D_{0} \) whenever \( \{ \} < 0 \), i.e.,

\[
\frac{f(r \mid H_{1})}{f(r \mid H_{0})} \begin{cases} \\
H_{1} & \text{H}_{1} \rangle = \frac{(C_{10} - C_{00})P(H_{0})}{(C_{01} - C_{11})P(H_{1})} \\
H_{0} & \text{H}_{0} \langle \end{cases} \\
\]

where \( \langle \) means “determine \( H_{1} \text{ if } \{ \} < \) is true”.
Define the likelihood ratio (LR), \( \Lambda(r) \)
\[
\Lambda(r) = \frac{f(r \mid H_1)}{f(r \mid H_0)}
\]
and the threshold \( \eta_{Bayes} \) is
\[
\eta_{Bayes} = \frac{(C_{10} - C_{00})P(H_0)}{(C_{01} - C_{11})P(H_1)}.
\]

The above argument tells us that the Bayes criterion leads us to a likelihood ratio test (LRT)
\[
\begin{array}{c|c}
H_1 & \Lambda(r) > \eta_{Bayes} \\
H_0 & \Lambda(r) < \eta_{Bayes}
\end{array}
\]
which is tantamount to saying that the decision regions
\[
D_0 = \{ r \mid \Lambda(r) < \eta_{Bayes} \}
\]
\[
D_1 = \{ r \mid \Lambda(r) > \eta_{Bayes} \}
\]
minimize the average cost.

In many cases, it is convenient to use the equivalent test
\[
\ln \Lambda(r) > \ln \eta_{Bayes}
\]
since \( \ln(\cdot) \) is monotonically increasing.
This test is commonly known as the log likelihood ratio test (LLRT).

\[
<Ex*> \quad H_1 : r(\mu) = m_1 + n(\mu) \quad H_0 : r(\mu) = n(\mu)
\]
\[
\begin{array}{ccc}
\vec{m_1} & \oplus & r(\mu) \\
& \oplus & n(\mu) \\
& & 0 \oplus r(\mu)
\end{array}
\]
where (1) \( \vec{1} \) is an N-dimensional all-one vector
(2) \( r(\mu) \) and \( n(\mu) \) are both N-dimensional
(3) \( n(\mu) \) contains iid Gaussian rv’s, which have zero mean and variance \( \sigma^2 \).
Now,

\[ \Lambda(r) = \frac{f(r \mid H_1)}{f(r \mid H_0)} = \frac{\prod_{i=1}^{N} f_{n_i}(r_i - m)}{\prod_{i=1}^{N} f_{n_i}(r_i)} = \frac{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(r_i - m)^2}{2\sigma^2}}}{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{r_i^2}{2\sigma^2}}} = \prod_{i=1}^{N} e^{-\frac{1}{2\sigma^2}(m^2 - 2r_i m)} \]

\[ \Rightarrow \ln \Lambda(r) = -\frac{Nm^2}{2\sigma^2} + \frac{m}{\sigma^2} \sum_{i=1}^{N} r_i \]

\[ \begin{array}{c|c}
H_1 & \ln \eta_{\text{Bayes}} \\
\hline
\sum_{i=1}^{N} r_i & \alpha \\
\hline
H_0 & < \\
\end{array} \]

where \( \alpha = \frac{\sigma^2}{m} \ln \eta_{\text{Bayes}} + \frac{Nm}{2} \)

Notes:

(1) Knowing \( \ell(r(\mu)) = \sum_{i=1}^{N} r_i(\mu) \) is just as good as knowing \( r \) for the purpose of decision optimization. This \( \ell(\mu) \) is called the sufficient statistic for this Bayes decision problem.

(2) In this example, we have a sufficient statistic which is a linear function of \( \{r_i(\mu)\}_{i=1}^{N} \).