* Receiver Operating Characteristic (ROC)

- The performance of an LRT can be completely characterized by the probability pair \((P_F, P_D)\).

The figure is commonly known as the ROC for the test.

**Ex. (from Ex *)**

\[
\ell(r) = \frac{1}{\sqrt{N\sigma}} \sum_{i=1}^{N} r_{i} \begin{cases} < \sqrt{N\eta/m} & \text{if } H_{i} \leq H_{0} \\ > \sqrt{N\eta/m} & \text{if } H_{i} > H_{0} \end{cases}
\]

Since \(\ell(r(\mu))\) is Gaussian with \(E\{\ell(r(\mu))|H_{0}\} = 0\), \(E\{\ell(r(\mu))|H_{1}\} = \frac{\sqrt{N\eta/m}}{\sigma}\),

\[
Var\{\ell(r(\mu))|H_{0}\} = 1, \quad \text{and} \quad Var\{\ell(r(\mu))|H_{1}\} = 1
\]

\[
P_F = \int_{\frac{\ln\eta}{d} + \frac{d}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = Q\left(\frac{\ln\eta}{d} + \frac{d}{2}\right)
\]

\[
P_D = \int_{\frac{\ln\eta}{d} - \frac{d}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-d)^2}{2}} dx = Q\left(\frac{\ln\eta}{d} - \frac{d}{2}\right)
\]

where \(d = \frac{\sqrt{N\eta/m}}{\sigma} = E\{\ell(r(\mu))|H_{1}\} - E\{\ell(r(\mu))|H_{0}\}\),

and \(Q(x)\) is the Gaussian tail integral \(Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy\).
Note that \( P_D = Q(Q^{-1}(P_F) - d) \) with \( Q^{-1} \) being the inverse of \( Q \).

- Bounds on \( Q(x) \)

\( Q(x) \) is a frequently-used function in digital communications. It is well-tabulated. For analytical convenience, it is useful to discuss its bounds, as follow:

\[
\frac{1}{\sqrt{2\pi}x}(1 - \frac{1}{x^2})e^{-\frac{x^2}{2}} < Q(x) < \frac{1}{\sqrt{2\pi}x} e^{-\frac{x^2}{2}}, \quad x > 0
\]

- Defn: An LRT with \( \Lambda(r(\mu)) \) being a continuous random variable is called a continuous LRT.

- Properties of ROC:

1. All continuous LRT’s have ROC’s that are concave downward.
2. All continuous LRT’s have ROC’s that are equal to or above the line $P_D = P_F$.

3. The slope of a curve in an ROC at a particular point is equal to the value of $\eta$ to achieve the $(P_D, P_F)$ of that point.

**pf:**

$$P_D = \int_{\eta}^{\infty} f(\Lambda \mid H_1)d\Lambda \quad\quad P_F = \int_{\eta}^{\infty} f(\Lambda \mid H_0)d\Lambda$$

**a.** Now,

$$\frac{dP_D}{dP_F} = \frac{dP_D / d\eta}{dP_F / d\eta} = -\frac{f(\Lambda = \eta \mid H_1)}{f(\Lambda = \eta \mid H_0)}$$

**b.** Next, defining $\Omega(r) = \{r \mid \Lambda(r) > \eta\} = \{r \mid \frac{f(r \mid H_1)}{f(r \mid H_0)} > \eta\}$

we have $P_D(\eta) = \Pr\{\Lambda(\mu) > \eta \mid H_1\}$

$$= \int_{\Omega(r)} f(r \mid H_1)dr$$

$$= \int_{\Omega(r)} \Lambda(x) f(r \mid H_0)dr$$

By use of the transformation $X(\mu) = \Lambda(\mu) \quad (f(X \mid H_0)dX = f(r \mid H_0)dr)$,

we have $P_D(\eta) = \int_{\eta}^{\infty} X f(X \mid H_0)dX$

$$\Rightarrow \frac{dP_D(\eta)}{d\eta} = -\eta f(\Lambda = \eta \mid H_0) = -\eta f(\Lambda = \eta \mid H_0)$$

**c.** Finally,

$$\frac{dP_D}{dP_F} = \frac{dP_D / d\eta}{dP_F / d\eta} = -\frac{f(\Lambda = \eta \mid H_0)}{f(\Lambda = \eta \mid H_0)} = \eta$$

**Q.E.D.**
§M-ary Decision

Consider

* Bayes Criterion

- Defns: 1. \( c_{ij} \equiv \text{the cost that } H_j \text{ is sent and } H_i \text{ is detected.} \)

  2. \( c \equiv \text{average cost} \)

\[
= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \Pr\{H_j \text{ is sent and } H_i \text{ is detected}\} c_{ij}
\]

\[
= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P(H_j) c_{ij} \int_{D_i} f(r \mid H_j) dr
\]

Since \( D = \bigcup_{k=0}^{M-1} D_k \) and \( D_k \)'s are disjoint,

\[
- c = \sum_{i,j=0}^{M-1} P(H_j) c_{ij} \int_{D_i} f(r \mid H_j) dr + \sum_{j=0}^{M-1} P(H_j) c_{ij} \int_{D_j} f(r \mid H_j) dr
\]

\[
= \sum_{i,j=0}^{M-1} P(H_j) (c_{ij} - c_{jj}) \int_{D_i} f(r \mid H_j) dr + \sum_{j=0}^{M-1} P(H_j) c_{jj}
\]

\[
= \sum_{j=0}^{M-1} P(H_j) c_{jj} + \sum_{i=0}^{M-1} \int_{D_i} \sum_{j=0}^{M-1} P(H_j) (c_{ij} - c_{jj}) f(r \mid H_j) dr
\]

- \( c \) is minimized if we choose \( D_i \) whenever
\[ I_i(r) = \sum_{j=0}^{M-1} P(H_j)(c_{ij} - c_{jj})f(r \mid H_j) \]

is minimum among all \( M \) possible values. This decision rule is stated as

\[
\text{Determine } H_i \text{ if } I_i(r) = \min_k I_k(r) \quad \text{...........(*)}
\]

- Define \( \Lambda_i(r) = \frac{f(r \mid H_i)}{f(r \mid H_0)} \), \( i = 0, 1, ..., M - 1 \)

It follows from (*) that the LRT for this \( M \)-ary decision is:

\[
\text{determine } H_i \text{ if } \sum_{j=0}^{M-1} P(H_j)(c_{ij} - c_{jj})\Lambda_j(r) = \min_k \sum_{j=0}^{M-1} P(H_j)(c_{ij} - c_{jj})\Lambda_j(r) \quad \text{...........(+)}
\]

where we let \( \Lambda_0(r) = 1 \) by default.

Notes:

1. (+) says that \( \Lambda_1(r), ..., \Lambda_{M-1}(r) \) are sufficient for this \( M \)-ary decision problem.

![Diagram](image)

This implies that “at most \((M-1)\) dimensions are necessary for constructing the decision space of an \( M \)-ary decision problem.”

2. At most \((M-1)\) sufficient statistics are necessary to contain all the information for making a decision.

3. When \( c_{jj} = 0 \), \( \forall j \) (no cost for correct decision) and \( c_{ij} = 1 \), \( \forall i \neq j \)
(equal unit cost for incorrect decision), we have
\[
- \ c = \sum_{j=0}^{M-1} \sum_{i=0}^{M-1} P(H_j) \int_{D_i} f(r \mid H_j) \, dr
\]
\[
\Pr \{ H_j \text{ is sent and } H_i \text{ is decided} \}
\]
, i.e., the error probability, and the decision rule
\[
\sum_{j=0}^{M-1} P(H_j) \Lambda_j (r) = \min_{k \neq i} \sum_{j=0}^{M-1} P(H_j) \Lambda_j (r)
\] ........................(x)
leads us to an M-ary MEP test.
Now, dividing both sides of (x) by \( P(H_0) \) and defining that the a posteriori Probability ratio
\[
\Lambda_j^* (r) = \frac{P(H_i)}{P(H_0)} \Lambda_j (r) = \frac{P(H_i) f(r \mid H_i)}{P(H_0) f(r \mid H_0)}
\]
\[
= \frac{P(H_i \mid r) f(r)}{P(H_0 \mid r) f(r)}
\]
\[
= \frac{P(H_i \mid r)}{P(H_0 \mid r)}
\]
The MEP test is equivalent to “determining \( H_i \) if
\[
\sum_{j=0}^{M-1} \Lambda_j^* (r) = \min_{k \neq i} \sum_{j=0}^{M-1} \Lambda_j^* (r)
\]
\[
\Rightarrow \ \frac{1 - P(H_i \mid r)}{P(H_0 \mid r)} = \min_{k \neq i} \frac{1 - P(H_k \mid r)}{P(H_0 \mid r)}
\]
\[
\Rightarrow \ P(H_i \mid r) = \max_k P(H_k \mid r)
\]
which is a “maximum a posteriori probability” (MAP) test.

4. Question: can you find the extensions to M-ary minimax test and NP test?
Parameter Estimation Theory

Based on the observed $r$, the estimation rule is to extract the parameter $\hat{a}$ being sent from the parameter space through the probabilistic mapping.

- Difference between hypothetical decision and parameter estimation is “hypothetic decision is to classify $r(\mu)$ out of $M$ finite possible classes,” while parameter estimation is “to extract parameter out of $r(\mu)$, which takes on value out of (usually) an infinitely countable set or an infinite set.”

Ex: The problem that distinguishes two hypotheses

$$H_0 : r(\mu) = n(\mu) \quad H_1 : r(\mu) = a_1 + n(\mu)$$

is a decision problem.

The problem that determines $a$ out of $r(\mu) = a_1 + n(\mu)$ probabilistic model is a parameter estimation problem.

- The parameter $a$ to be estimated may be “random” or “nonrandom but unknown.”

§ Random Parameter Estimation (Considering $a(\mu) = a(\mu)$)

- Here, $a(\mu)$ is random and assumed of continuous type (in the real line).

- Random estimation is commonly known as Bayes estimation.

* Bayes Estimation
• Defns: 1. \( a_e \equiv \hat{a}(r) - a \) is the error of estimate \( \hat{a}(r) \), given that \( a(\mu) = a \) is sent.

2. \( c(a_e) \) is the cost associated with the estimation error \( a_e \).

• The cost function \( c(a_e) \) is commonly attributed with two properties:

1. \( c(a_e) = c(-a_e) \) (symmetric)

2. \( 0 = c(0) \leq c(a_e) \)

• Three types of cost functions are considered:

  a. \( c(a_e) = a_e^2 \) (mean-square (MS) error cost)

  b. \( c(a_e) = |a_e| \) (absolute error cost)

  c. \( c(a_e) = \begin{cases} 0, & \text{if } |a_e| < \frac{\Delta}{2} \\ 1, & \text{otherwise} \end{cases} \) (uniform cost)

• The Bayes estimation is to find an estimate \( \hat{a}(r) \) that minimizes the \( E\{c(a_e(\mu))\} \), provided with a tractable cost function.

Note: 1. Bayes estimation assumes that the density \( f(a) \) is known a priori.

2. When \( f(a) \) is unknown, a minimax estimation, which finds \( \hat{a} \) that minimizes \( \max_{f(a)} E\{c(a_e)\} \), can be used. (not treated below)

Now, \( E\{c(a_e(\mu))\} = E\{c[a(r(\mu)) - a(\mu)]\} \)

\[
= \int_{-\infty}^{\infty} \int_{D} c[\hat{a}(r) - a] f(a, r) dr da
\]

\[
= \int_{D} \int_{-\infty}^{\infty} c[\hat{a}(r) - a] f(a | r) da f(r) dr
\]
The root of \( \frac{\partial}{\partial \hat{a}} E\{c(a_\sigma(\mu))\} = 0 \) may minimize \\
\( E\{c(a_\sigma(\mu))\} \) because the above integrand is \\
nonnegative. Furthermore, since only the inner \\
integral depends on \( \hat{a} \), Bayes estimation is \\
equivalent to finding the root of \\

\[
\frac{\partial}{\partial \hat{a}} \int_{-\infty}^{\infty} c[\hat{a}(r) - a] f(a \mid r)da = 0
\]

\\[\text{...............}(\#)\\]

- For the mean-square (MS) error function, \\

\[
(\#) \quad \Rightarrow -2 \int_{-\infty}^{\infty} a f(a \mid r)da + 2 \hat{a}_{MS}(r) \int_{-\infty}^{\infty} f(a \mid r)da = 0
\]

\[
\Rightarrow \hat{a}_{MS}(r) = \int_{-\infty}^{\infty} a f(a \mid r)da = E\{a(\mu) \mid r\}
\]

i.e., the MS estimate is the mean of the \textit{a posteriori} probability density.

- For the absolute (ABS) error function, \\

\[
(\#) \quad \Rightarrow \frac{\partial}{\partial \hat{a}} \int_{-\infty}^{\infty} |\hat{a} - a| f(a \mid r)da = 0
\]

\[
\Rightarrow \frac{\partial}{\partial \hat{a}} \int_{-\infty}^{\hat{a}} (\hat{a} - a)f(a \mid r)da + \frac{\partial}{\partial \hat{a}} \int_{\hat{a}}^{\infty} (a - \hat{a})f(a \mid r)da = 0
\]

\[
\Rightarrow \int_{-\infty}^{\hat{a}_{ABS}} f(a \mid r)da = \int_{\hat{a}_{ABS}}^{\infty} f(a \mid r)da = \frac{1}{2}
\]

i.e., the ABS estimate is the median of the \textit{a posteriori} probability
density.

- For the uniform (UNF) cost function,

The Bayes estimation has to minimize

\[
\hat{a}_{UNF} \text{ cost} + \frac{\Delta}{2}
\]

\[
1 - \int_{\hat{a}_{UNF} - \frac{\Delta}{2}}^{\hat{a}_{UNF} + \frac{\Delta}{2}} f(a \mid r) da
\]

When \( \Delta \) is arbitrarily close to zero (with \( \Delta > 0 \)), this implies that \( \hat{a}_{UNF}(r) \) should be the value that maximizes \( f(a \mid r) \). That is, the UNF estimate is the value of \( a \) that yields the maximum \( a \) posteriori probability density value.

Note: Such an estimate is called a maximum \( a \) posteriori probability (MAP) estimate.