Detection & Estimation 304-621

1. Introduction

→ Extraction of information from noisy signals.

Detection : deciding which of a finite set of possibilities is occuring
(ex.: does a noisy signal correspond to 0 or 1?)

Estimation : Forming an opinion about the numerical value of a quantity which cannot be observed directly (ex.: target estim. distance by radar: bounce FM signal & measure time to estim. distance)

2. Review & Notation

Random Variables : capital letters , Realizations : lowercase letters

$F_X(x)$ denotes "distribution function" : a.k.a. cumulative distribution function

$f_X(x)$ denotes "probability density function" for continuous RV.

with $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$

$\{Pr(X=x_i)\}$ denote probability mass function associated with a discrete RV.

with $F_X(x) = \sum_{u=-\infty}^{x} Pr(X=u)$

Unified Notation

$$F_X(x) = \int_{-\infty}^{x} dF_X(y) = \begin{cases} \text{continuous}: & \int_{-\infty}^{x} f_X(y)\,dy \\ \text{discrete}: & \sum_{y\leq x} Pr\{X=y\} \end{cases}$$

Random vectors denoted by $\underline{X}$ (always row vectors)

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \text{ each entry is a RV.}$$

x_i all continuous $\to$ $\underline{x}$ is called continuous etc...

Joint (multivariate) distribution function of $\underline{x}$ is:

$$F_{\underline{x}}(\underline{x}) = Pr\{x_1 \le x_1, \; x_2 \le x_2 \; \dots\}$$

$\boxed{\text{Properties}}$ $f_{\underline{x}}(x_1, \dots x_n)$ "JPDF" s.t. $F_{\underline{x}}(\underline{x}) = \int_{-\infty}^{\underline{x}} \cdots \int f_{\underline{x}}(x_1, \dots x_n) dx_n \dots dx_1$

★ statistical independence of components $\Leftrightarrow$

$$F_{\underline{x}}(\underline{x}) = Pr\{x_1 \le x_1\} \cdot Pr\{x_2 \le x_2\} \cdot \dots$$

① Marginal DF of x_i $\triangleq F_{\underline{x}}(\infty, \infty, \dots, x_i, \dots \infty, \infty)$

② $F_{\underline{x}}\{x_1, x_2, \dots, -\infty, \dots x_n\} = 0$

③ $F_{\underline{x}}\{\infty, \infty, \dots, \infty\} = 1$

④ $0 \le F_{\underline{x}}(\underline{x}) \le 1$

$\boxed{\text{Summary}}$

Continuous $\underline{x}$	Discrete $\underline{x}$
$F_{\underline{x}}(\underline{x}) = \int_{-\infty}^{\underline{x}} f_{\underline{x}}(y_1, \dots y_n) d\underline{y}$ $= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots f_{\underline{x}}(y_1, \dots y_n) dy_1 dy_2 \dots$	$F_{\underline{x}}(\underline{x}) = \sum_{\underline{y}=0}^{\underline{x}} Pr(x_1 = y_1, \dots x_n = y_n)$
$F_{\underline{x}}(\underline{x}) = Pr\{x_1 \le x_1, x_2 \le x_2 \dots\}$	JPMF $f_{\underline{x}}(\underline{x}) = \{Pr(x_1 = x_1, \dots x_n = x_n)\}$ ($n \times n$ array of probs)

Unified Notation

$$F_{\underline{x}}(\underline{x}) = \int_{-\infty}^{\underline{x}} dF_{\underline{x}}(\underline{y}) = \begin{cases} \int_{-\infty}^{\underline{x}} f_{\underline{x}}(y_1, \dots y_n) \, d\underline{y}, & \text{continuous} \\ \sum_{\underline{y}=0}^{\underline{x}} Pr(x_1 = y_1, \dots x_n = y_n), & \text{discrete} \end{cases}$$

Let Z be a subset of $\mathbb{R}^n$ for which a probability measure can be defined; Z is then a Borel subset. Then,

$$P\{\underline{x} \in Z\} = \int_Z dF_{\underline{x}}(\underline{x})$$

Conditional probability

$$P(X|Y) = \frac{P(X, Y)}{P(Y)}$$

$$\Downarrow$$

$$F_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = P[\underline{X} \leq \underline{x} \,|\, \underline{Y} = \underline{y}]$$

$$= \int_{-\infty}^{\underline{x}} f_{\underline{X}|\underline{Y}}(\underline{u}|\underline{y})\, d\underline{u}$$

$$\Downarrow$$

$$f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{f_{\underline{X}\underline{Y}}(\underline{x}, \underline{y})}{f_{\underline{Y}}(\underline{y})}$$

$\underline{x}, \underline{y}$ independent iff $f_{\underline{x}|\underline{y}}(\underline{x}|\underline{y}) = f_{\underline{x}}(\underline{x})$

Means, moments & characteristic functions

given Random vector $\underline{x}$

Random variable $g(\underline{x})$

we write $E(g(\underline{x})) = \int g(\underline{x})\, dF_{\underline{x}}(\underline{x}) = \begin{cases} \int g(\underline{x}) f_{\underline{x}}(\underline{x}) d\underline{x} \\ \sum_{\underline{x}} g(\underline{x}) P(\underline{X} = \underline{x}) \end{cases}$

n^{th} order moment of a RV y is:

$$E(y^n) = \int y^n dF_y(y)$$

$n=1 \rightarrow$ mean : $E(y)$

$n=2 \rightarrow$ variance : $E((y-\mu_y)^2)$

mean of a vector $\rightarrow$ mean of its components.

correlation matrix of $x \rightarrow R_{\underline{x}} = E\{\underline{x}\,\underline{x}^T\}$

$R_{\underline{x}\,i,k} = E(y_i\, y_k)$

covariance matrix $\Rightarrow C_x = E((\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})^T)$

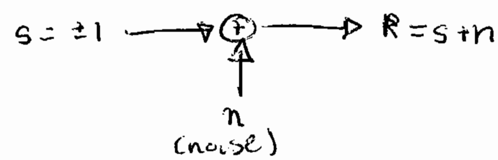
3 DETECTION THEORY (HYPOTHESIS TESTING)

(read ch. 2)

3.1 Preliminaries

Model: Decision rule ~~on~~ between several Hypothesis, based on one realization
(several realizations are treated as a vectorial observation)

$\boxed{\text{Ex}}$



Observe R, decide on s = 1 v.s s = −1

Let n be a discrete RV with pmf: $f_n(n) = \begin{cases} 1/3, & n = 1 \\ 1/2, & n = 0 \\ 1/6, & n = -1 \end{cases}$

Formally:

$$\left.\begin{aligned} H_0 &: R = 1 + n \\ H_1 &: R = -1 + n \end{aligned}\right\} \to \text{correspond to different prob. laws for } R.$$

$$H_1 \to f(r|H_1) = \begin{cases} 1/3, & 2 = n \\ 1/2, & 1 = n \\ 1/6, & 0 = n \end{cases}$$

$$H_0 \to f(r|H_0) = \begin{cases} 1/3, & 0 = n \\ 1/2, & -1 = n \\ 1/6, & -2 = n \end{cases}$$

p(noise = −1 & s = 1

$\Downarrow$

observation space: $\{-2, -1, 0, 1, 2\} = Z$

a Decision rule: $r \in \{-2, -1, 0\} \to H_0$ is chosen
$r \in \{1, 2\} \to H_1$ is chosen

$\to$ a "partition" of the observation space.
with properties: $Z_0 \cup Z_1 = Z$
$Z_0 \cap Z_1 = \phi$

Notation a partition corresponding to a decision rule can be described as a set of sets:

$$\delta = \{Z_0, Z_1, \dots Z_n\}$$

when several partitions are used:

$$\delta^0 = \{Z_0^0, Z_1^0, \dots Z_n^0\}$$
$$\delta^1 = \{Z_0^1, Z_1^1, \dots Z_n^1\}$$
etc...

$\boxed{\text{Def}}$ Randomized decision rule for testing M hypothesis is a set of (simple) decision rules $\delta_R = \{\delta^0, \delta^1, \dots \delta^{M-1}\}$ and the associated discrete probability law $f_R = \{p_0, p_1, \dots p_{M-1}\}$. The decision rule consists of two steps: 1) according to f_R, choose δ^i

2) use δ^i to make the decision.

$\Rightarrow$ simple decision rule is the special case where f_R only has one non-zero element. $\Rightarrow$ δ^j always selected.

Randomized decision rule can be represented as

$$\delta^I = \{Z_0^I, Z_1^I, \dots Z_{M-1}^I\}$$

where I is a discrete RV with probability law $\Pr\{I = i\} = p_i$, $i = 0, 1, \dots K-1$

$\boxed{\text{Performance criterion for decision rules}}$

$\hookrightarrow$ based on probability of error

$\boxed{\text{Def}}$ Probability of error associated with H_m is

$$P_e(H_m) = \text{prob}(\overset{\text{do not}}{\text{chose}}\ H_m \mid H_m \text{ is true})$$

$\boxed{Ex}$ (see previous example)

$$P_e(H_1) = P(H_0 | H_1)$$
$$= 1/6$$

$$P_e(H_0) = 0$$

For a given randomized decision rule $\delta^I = \{ Z_0^I, Z_1^I, \dots \}$, the prob. of error when $I = i$ is

$$P_e(H_m | I=i) = \int_{\bar{Z}_m^i} dF_{\underline{R}|H}(\underline{r} | H_m)$$

complement

$$= 1 - \int_{Z_m^i} dF_{\underline{R}|H}(r | H_m)$$

$$= 1 - P_c(H_m | I = i)$$

the probability of error associated with H_m for the randomi decision rule δ^I is:

$$P_e(H_m) = E_I [P_e(H_m | I)]$$
$$= \sum_{i=0}^{k-1} P_i \, P_e(H_m | I=i)$$
$$= E_I \int_{\bar{Z}_m^I} dF_{\underline{R}|H}(\underline{r} | H_m)$$
$$= \sum_{i=0}^{k-1} \left(P_i \int_{\bar{Z}_m^i} dF_{\underline{R}|H}(\underline{r} | H_m) \right)$$
$$= E_I [1 - P_c(H_m | I)]$$
$$= 1 - E_I [P_c(H_m | I)]$$
$$= 1 - P_c(H_m)$$

$\boxed{\text{Define}}$ the <u>characteristic function of region Z_m^i</u> as

$$1_{Z_m^i} = \begin{cases} 1, & \underline{r} \in Z_m^i \\ 0, & \text{else} \end{cases}$$

Then:

$$P_e(H_m) = \sum_{i=0}^{K-1} p_i \left[1 - \int 1_{Z_m^i}(\underline{r})\, dF_{\underline{R}|H}(\underline{r}|H_m)\right]$$

$$= 1 - \sum_{i=0}^{K-1} p_i \, E\left[1_{Z_m^i}(\underline{R}) | H_m\right]$$

$$= 1 - E_I\left[E\left[1_{Z_m^i}(\underline{R}) | H_m, I\right]\right]$$

$\boxed{\text{Def}}$ <u>Pairwise probability of error</u>:

- Probability that when $I = i$ the decision is H_n when H_m is true:

$$P(H_n | H_m, I=i) = \int_{Z_n^i} dF_{\underline{R}|H}(\underline{r}|H_m)$$

- Probability that the decision is H_n when H_m is true:

$$P(H_n|H_m) = E_I\left[\int_{Z_n^I} dF_{\underline{R}|H}(\underline{r}|H_m)\right]$$

$$= \sum_{i=0}^{K-1} p_i \int_{Z_n^i} dF_{\underline{R}|H}(\underline{r}|H_m)$$

(i.e. all i's are still possible)

We also have:

$$\sum_{n=0}^{M-1} P(H_n|H_m, I=i) = \sum_{n=1}^{M} \int_{Z_n^i} dF_{\underline{R}|H}(\underline{r}|H_m)$$

$$= \int_{Z} dF_{\underline{R}|H}(\underline{r}|H_m)$$

$$= 1$$

3.2 Likelihood Ratio (LR)

Consider two Hypotheses:

$$H_0 \to F_{\underline{R}|H}(\underline{r}|H_0) \begin{cases} \xrightarrow{\text{cont}} f_{\underline{R}|H}(\underline{r}|H_0) \\ \xrightarrow{\text{discrete}} P_{\underline{R}|H}(\underline{r}|H_0) \end{cases}$$

$$H_1 \to F_{\underline{R}|H}(\underline{r}|H_1) \begin{cases} \xrightarrow{\text{cont}} f_{\underline{R}|H}(\underline{r}|H_1) \\ \xrightarrow{\text{disc}} P_{\underline{R}|H}(\underline{r}|H_1) \end{cases}$$

Likelihood Ratio defined as follows:

$$\Lambda_{1,0}(\underline{r}) = \frac{dF_{\underline{R}|H}(\underline{r}|H_1)}{dF_{\underline{R}|H}(\underline{r}|H_0)} \begin{cases} \xrightarrow{\text{cont}} \dfrac{f_{\underline{R}|H}(\underline{r}|H_1)}{f_{\underline{R}|H}(\underline{r}|H_0)} \\ \xrightarrow{\text{disc.}} \dfrac{P_{\underline{R}|H}(\underline{r}|H_1)}{P_{\underline{R}|H}(\underline{r}|H_0)} \end{cases}$$

Lemma 3.1

let $g(\underline{r})$ be a function of an observation $\underline{r}$ in a 2-Hypotheses testing problem (H_0 vs. H_1) with L.R. $\Lambda_{1,0}(\underline{r})$. Then we have:

$$E[g(\underline{R})\Lambda_{1,0}(\underline{R})|H_0] = E[g(\underline{R})|H_1]$$

Proof continuous case

$$E[g(\underline{R})\Lambda_{1,0}(\underline{R})|H_0] = \int g(\underline{r}) \frac{f_{\underline{R}|H}(\underline{r}|H_1)}{f_{\underline{R}|H}(\underline{r}|H_0)} f_{\underline{R}|H}(\underline{r}|H_0)\, d\underline{r}$$

$$= \int g(\underline{r})\, f_{\underline{R}|H}(\underline{r}|H_1)\, d\underline{r}$$

$$= E\{g(\underline{R})|H_1\} \qquad \square$$

discrete case:

replace integral by $\sum_{\underline{r}}$, f by p

L.R. decision rule

$$\Lambda_{1,0}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$ — threshold

→ a randomised decision rule.

Decision regions associated with the test

$$Z_1 = \{\underline{r} \text{ s.t. } \Lambda_{1,0}(\underline{r}) > \lambda\}$$

$$Z_0 = \{\underline{r} \text{ s.t. } \Lambda_{1,0}(\underline{r}) < \lambda\}$$

$$Z_{1,0} = \{\underline{r} \text{ s.t. } \Lambda_{1,0}(\underline{r}) = \lambda\}$$ → sometimes probability of this is zero → ignore $Z_{1,0}$. (when Λ is continuous in)

⇒ if $\underline{r} \in Z_1$ → decide H_1

$\underline{r} \in Z_0$ → decide H_0

$\underline{r} \in Z_{1,0}$ → decide H_1 with prob. "q", otherwise H_0.

Probability of error in terms of L.R.

→ i.e. do not choose H_0, when H_0 is true.

$$P_e(H_0) = q\,P_r[\underline{R} \in Z_1 \cup Z_{1,0} | H_0] + (1-q)\,P_r[\underline{R} \in Z_1 | H_0]$$

$$= q\,P_r[\underline{R} \in Z_1 | H_0] + q\,P_r[\underline{R} \in Z_{1,0} | H_0] + (1-q)\,P_r[\underline{R} \in Z_1 | H_0]$$

$$= P_r[\underline{R} \in Z_1 | H_0] + q\,P_r[\underline{R} \in Z_{1,0} | H_0]$$

$$= P_r[\Lambda_{1,0}(\underline{R}) > \lambda | H_0] + q\,P_r[\Lambda_{1,0}(\underline{R}) = \lambda | H_0]$$

$$P_e(H_1) = P_r[\Lambda_{1,0}(\underline{R}) < \lambda | H_1] + (1-q)\,P_r[\Lambda_{1,0}(\underline{R}) = \lambda | H_1]$$

3.3 Neyman-Pearson framework

Consider a 2 hypotheses testing problem (H_0, H_1).
Define two types of errors, and their error probs;

$$P_e(H_0) = P(H_1 | H_0) \quad , \quad P_e(H_1) = P(H_0 | H_1)$$

$P_e(H_0) = P_F \rightarrow$ "prob. of false alarm"
"prob. of type 1 error"
"size of the test" or "test level"

$P_e(H_1)$ $[1 - P_e(H_1)] \xrightarrow{=P_D}$ "prob. of detection"
$P_e(H_1) \rightarrow$ "prob of type 2 error"
$[1 - P_e(H_1)] \rightarrow$ "power of the test"

Main Problem

Find a decision rule that minimizes $P_e(H_1)$ given a constraint $P_e(H_0) \leq \alpha$
i.e. find the most powerful test given a bound on the test level.

Theorem 3.1

Consider a two hypotheses testing problem (H_0 vs. H_1) and a likelihood ratio decision rule specified by the following:

$$A_{1,0}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \quad , \quad \lambda > 0 \quad , \quad \underset{Z_{1,0}}{Z_0, Z_1} \text{ decision regions given}$$

$\& \ \underline{r} \in Z_{1,0} \rightarrow$ choose H_1 with prob. q_1

not sure what this means.?

Let the prob. of errors be $P_e(H_0) = \alpha_{LR}$, $P_e(H_1) = \beta_{LR}$
consider another decision rule employing regions $Z_1'^{c}, Z_0'^{c}$ with $P_0, P_1, \ldots P_{k-1}$ and prob of error $P_e(H_0) = \alpha'$, $P_e(H_1) = \beta'$

$\Rightarrow$ If $\beta' < \beta_{LR}$ then $\alpha' > \alpha_{LR}$

Lemma (to be used in proof of theorem 3.1)

Define the following sets:

$$Z_1^j = \begin{cases} Z_1 \cup Z_{1,0} & , j=1 \\ Z_1 & , j=0 \end{cases}$$

$$Z_0^j = \begin{cases} Z_0 & , j=1 \\ Z_0 \cup Z_{1,0} & , j=0 \end{cases}$$

Then for $j=0,1$ and $I=0,\ldots k-1$ we have:

$$\boxed{\left[1_{Z_1^j}(\underline{r}) - 1_{Z_1^{i}}(\underline{r}) \right] \left[\Lambda_{1,0}(\underline{r}) - \lambda \right] \geq 0}$$

Proof

If $\underline{r} \in Z_1$: $1_{Z_1^j}(\underline{r}) - 1_{Z_1^i}(\underline{r}) = 1 - (1 \text{ or } 0) \geq 0$

$\Lambda_{1,0} - \lambda > 0$ (by def of Z_1)

If $\underline{r} \in Z_0$: $1_{Z_1^j}(\underline{r}) - 1_{Z_1^i}(\underline{r}) = 0 - (1 \text{ or } 0) \leq 0$

$\Lambda_{1,0} - \lambda < 0$

If $\underline{r} \in Z_{1,0}$: $\Lambda_{1,0} - \lambda = 0$

$\rightarrow$ product $\left[1_{Z_1^j}(\underline{r}) - 1_{Z_1^i}(\underline{r}) \right] \cdot \left[\Lambda_{1,0}(\underline{r}) - \lambda \right]$ is always non-negative.

Note: $1_{Z_1^j}(\underline{r}) = \begin{cases} 1 , & \underline{r} \in Z_1^j \\ 0 , & \underline{r} \notin Z_1^j \end{cases}$ (recalled)

Proof of Theorem 3.1

From the lemma we know:
(linear comb. of non-neg. values)

$$\sum_{j=0}^{1}\sum_{i=0}^{k-1} q_j p_i \, E\Big[\big[1_{z_j}(\underline{R}) - 1_{z_i'^i}(\underline{R})\big]\big[\Lambda_{1,0}(\underline{R}) - \lambda\big] \,\Big|\, H_0\Big] \geq 0$$

$$\sum_{j=0}^{1} q_j \cdot E\left[1_{z_1^j}(\underline{R})\Lambda_{1,0}(\underline{R})|H_0\right] - \sum_{i=0}^{k-1} p_i \, E\left[1_{z_1'^i}(\underline{R})\Lambda_{1,0}(\underline{R})|H_0\right]$$
$$-\lambda\sum_{j=0}^{1} q_j \cdot E\left[1_{z_1^j}(\underline{R})|H_0\right] + \lambda\sum_{i=0}^{k-1} p_i \, E\left[1_{z_1'^i}(\underline{R})|H_0\right] \geq 0$$

prob of errors.

Using lemma 3.1 on first 2 terms:

$$\sum_{j=0}^{1} q_j E\left[1_{z_1^j}(\underline{R})\Lambda_{1,0}(\underline{R})|H_0\right] = \sum_{j=0}^{1} q_j E\left[1_{z_1^j}(\underline{R})|H_1\right] = 1-\beta_{LR}$$

$$\sum_{i=0}^{k-1} p_i E\left[1_{z_1'^i}(\underline{R})\Lambda_{1,0}(\underline{R})|H_0\right] = \sum_{i=0}^{k-1} q_j \; E\left[1_{z_1'^i}(\underline{R}|H_1)\right] = 1-\beta'$$

Other two terms written directly as:

$$\lambda \sum_{j=0}^{1} q_j E\left[1_{z_1^j}(\underline{R})|H_0\right] = \lambda\cdot\alpha_{LR}$$
$$\lambda \sum_{i=0}^{k-1} p_i \, E\left[1_{z_1'^i}(\underline{R})|H_0\right] = \lambda\cdot\alpha'$$

Substituting yields:

$$(1-\beta_{LR}) - (1-\beta') - \lambda\alpha_{LR} + \lambda\alpha' \geq 0$$
$$-\beta_{LR} + \beta' + \lambda(\alpha' - \alpha_{LR}) \geq 0$$
$$\lambda(\alpha' - \alpha_{LR}) \geq \beta_{LR} - \beta'$$

$\rightarrow$ if $\beta_{LR} > \beta'$ then $\alpha' > \alpha_{LR}$

□

Going back to Neyman-Pearson objective:

"Find a decision rule that minimizes $P_e(H_1)$ under the constraint $P_e(H_0) \leq \alpha$."

- Theorem 3.1 implies that a likelihood ratio decision rule with $P_e(H_0) = \alpha$ is a solution (since any other decision rule with $P_e(H_0) < \alpha$ would lead to an increase in $P_e(H_1)$)

Solving $P_e(H_0) = \alpha$ gives:

$$P_e(H_0) = P_e[\Lambda_{1,0}(\underline{R}) > \lambda | H_0] + q_1 P_e[\Lambda_{1,0}(\underline{R}) = \lambda | H_0]$$

→ can adjust $P_e(H_0)$ through two parameters: λ & q_1

→ q_1 not relevant in case of continuous $\Lambda_{1,0}(\underline{R})$ (second term is zero).

Left with:

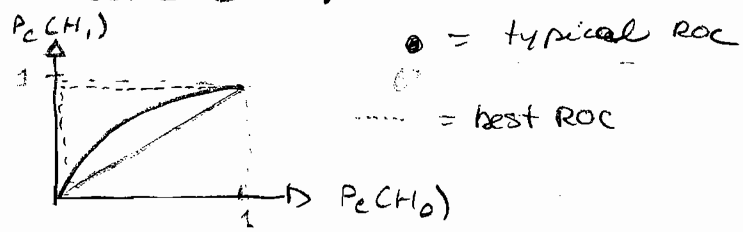
$$P_e(H_0) = P_e[\Lambda_{1,0}(\underline{R}) > \lambda | H_0] = \alpha$$

Note: for Λ discrete, may need randomized decision rule for certain values of α

How do we compare decision rules under Neyman-Pearson criterion?

- Performance given by $P_e(H_0)$ and $P_e(H_1)$
- But, traditionally we consider $P_e(H_0)$ and $P_c(H_1) = 1 - P_e(H_1)$
  - $P_c(H_1)$ and $P_e(H_0)$ are related through the threshold.
  - A graph of $P_c(H_1)$ v.s. $P_e(H_0)$ is called the Receiver Operating Characteristic (ROC)

Typical ROC:



- provides an indication of how much a decrease in false alarm decreases the detection probability & vice-versa.

$\boxed{\text{Ex 3.3.1}}$

Consider a 2-HT problem:

$$H_1: r_i = s + \mu_i, \quad i = 1, \dots, N$$
$$H_0: r_i = \mu_i, \quad i = 1, \dots, N$$

μ_i are i.i.d Gaussian realization with $N(0, \sigma^2)$
$s > 0$ is a known constant

Define $\underline{r} = [r_1 \ r_2 \ \cdots \ r_N]^T$ $\quad s = 1$

$$f(\underline{r}|H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r_i - s)^2}{2\sigma^2}}$$

$$f(\underline{r}|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{r_i^2}{2\sigma^2}}$$

$$\rightarrow \Lambda_{1,0}(\underline{r}) = \frac{f(\underline{r}|H_1)}{f(\underline{r}|H_0)} = \prod_{i=1}^{N} e^{\frac{2r_i s - s^2}{2\sigma^2}}$$

$$= e^{\frac{s}{\sigma^2}\sum_{i=1}^{N} r_i - N\frac{s^2}{2\sigma^2}}$$

$$= e^{\frac{Ns}{\sigma^2}\bar{r} - N\frac{s}{2\sigma^2}}$$

where $\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r_i$
"sample mean"

LR decision rule:

$$\Lambda_{1,0}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \quad \rightarrow \quad \ln \Lambda_{1,0}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \ln \lambda$$

$$\boxed{\bar{r} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma^2 \ln \lambda}{Ns} + 1/2}$$

→ the sample mean is a "sufficient statistics"
since it contains all the information from the received signal needed to make a decision.

→ $\bar{r} \sim N(s, \frac{\sigma^2}{N})$ under H_1, $\bar{r} \sim N(0, \frac{\sigma^2}{N})$ under H_0

3.2 Likelihood Ratio (LR)

Consider two Hypotheses:

$$H_0 \to F_{\underline{R}|H}(\underline{r}|H_0) \begin{cases} \xrightarrow{\text{cont}} f_{\underline{R}|H}(\underline{r}|H_0) \\ \xrightarrow{\text{discrete}} P_{\underline{R}|H}(\underline{r}|H_0) \end{cases}$$

$$H_1 \to F_{\underline{R}|H}(\underline{r}|H_1) \begin{cases} \xrightarrow{\text{cont}} f_{\underline{R}|H}(\underline{r}|H_1) \\ \xrightarrow{\text{disc}} P_{\underline{R}|H}(\underline{r}|H_1) \end{cases}$$

Likelihood Ratio defined as follows:

$$\Lambda_{1,0}(\underline{r}) = \frac{dF_{\underline{R}|H}(\underline{r}|H_1)}{dF_{\underline{R}|H}(\underline{r}|H_0)} \begin{cases} \xrightarrow{\text{cont}} \dfrac{f_{\underline{R}|H}(\underline{r}|H_1)}{f_{\underline{R}|H}(\underline{r}|H_0)} \\ \xrightarrow{\text{disc.}} \dfrac{P_{\underline{R}|H}(\underline{r}|H_1)}{P_{\underline{R}|H}(\underline{r}|H_0)} \end{cases}$$

Lemma 3.1

Let $g(\underline{r})$ be a function of an observation $\underline{r}$ in a 2-Hypotheses testing problem (H_0 vs. H_1) with L.R. $\Lambda_{1,0}(\underline{r})$. Then we have:

$$E[g(\underline{R})\Lambda_{1,0}(\underline{R})|H_0] = E[g(\underline{R})|H_1]$$

Proof continuous case

$$E[g(\underline{R})\Lambda_{1,0}(\underline{R})|H_0] = \int g(\underline{r}) \frac{f_{\underline{R}|H}(\underline{r}|H_1)}{f_{\underline{R}|H}(\underline{r}|H_0)} f_{\underline{R}|H}(\underline{r}|H_0)\, d\underline{r}$$

$$= \int g(\underline{r})\, f_{\underline{R}|H}(\underline{r}|H_1)\, d\underline{r}$$

$$= E\{g(\underline{R})|H_1\} \qquad \square$$

discrete case:

replace integral by $\sum_{\underline{r}}$, f by p

Prob. of error obtained:

$$P_e(H_0) = P_r\left[x > \frac{\sigma^2 \ln\lambda}{NS} + 1/2 \;\middle|\; H_0\right]$$

$$= P_r\left[\underbrace{\frac{\sqrt{N}}{\sigma}x}_{\sim N(0,1)} > \frac{\sigma \ln\lambda}{\sqrt{N}\,S} + \frac{\sqrt{N}\,S}{2\sigma} \;\middle|\; H_0\right]$$

$$\Rightarrow P_e(H_0) = Q\left(\frac{\sigma \ln\lambda}{\sqrt{N}\,S} + \frac{\sqrt{N}\,S}{2\sigma}\right)$$

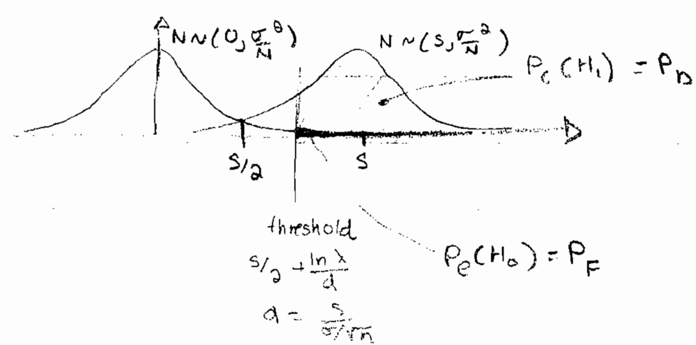
What is the probability of detection?

$$P_e(H_1) = P_r\left[x > \frac{\sigma^2 \ln\lambda}{NS} + 1/2 \;\middle|\; H_1\right]$$

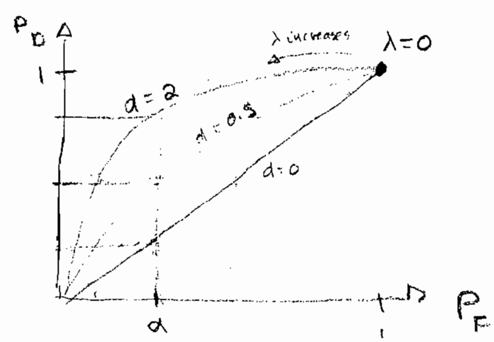
$$= P_r\left[\frac{\sqrt{N}(x-S)}{\sigma} > \frac{\sigma \ln\lambda}{\sqrt{N}\,S} + \frac{S\sqrt{N}}{2\,2} \;\middle|\; H_1\right]$$

$$= Q\left(\frac{\sigma \ln\lambda}{\sqrt{N}\,S} + \frac{S}{2}\frac{\sqrt{N}}{2}\right)$$

What is the distribution of x under H_0 and H_1?

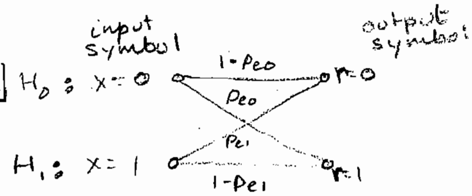


Plot P_D v.s. P_F : see hand out "3.19"



Notes
- large d → better tradeoff
- $\lambda\uparrow$ → P_D and P_F ↓ (good for P_F, bad for P_D.)

$\boxed{\text{Ex. 3.3.2}}$



<u>Assume</u>

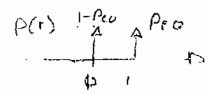
$\star$ $P_{e0} + P_{e1} < 1$

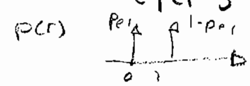
$0 < P_{e0} < 1$

$0 < P_{e1} < 1$

<u>Find L.R.</u>

$$p(r|H_0) = \begin{cases} 1-P_{e0}, & r=0 \\ P_{e0}, & r=1 \end{cases} \qquad p(r|H_1) = \begin{cases} P_{e1}, & r=0 \\ 1-P_{e1}, & r=1 \end{cases}$$





$$\Lambda_{1,0} = \frac{p(r|H_1)}{p(r|H_0)} = \begin{cases} \frac{P_{e1}}{1-P_{e0}} = a, & r=0 \\ \frac{1-P_{e1}}{P_{e0}} = b, & r=1 \end{cases}$$

$\Rightarrow$ discrete L.R.

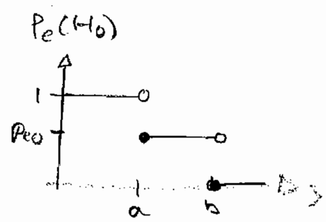
$\Rightarrow$ $P_{e0}+P_{e1} < 1 \rightarrow a < b$

<u>Find threshold</u>

$$P_{\Lambda_{1,0}}(a|H_0) = P(r=0|H_0) = 1-P_{e0}$$

$$P_{\Lambda_{1,0}}(b|H_0) = P_{e0}$$

$$P_e(H_0) = Pr\{\Lambda_{1,0}(R) > \lambda | H_0\} = \begin{cases} 1, & \lambda < a \\ P_{e0}, & a \le \lambda < b \\ 0, & b \le \lambda \end{cases}$$



$\rightarrow$ can only use this approach to find Neyman-Pearson decision rule for $\alpha = 0, 1$ or P_{e0}.

$\rightarrow$ otherwise, must use randomized decision rule.

i.e. instead of: $\Lambda_{1,0}(r) > \lambda \rightarrow H_1$
$\Lambda_{1,0}(r) \le \lambda \rightarrow H_0$

Use something like: $\Lambda_{1,0}(r) > \lambda \rightarrow H_1$
$\Lambda_{1,0}(r) < \lambda \rightarrow H_0$
$\Lambda_{1,0}(r) = \lambda \rightarrow H_1$ with prob q_1,

Using this randomized test

$$P_e(H_0) = Pr\{\Lambda_{1,0}(R) > \lambda \mid H_0\} + q_1 \cdot Pr\{\Lambda_{1,0}(R) = \lambda \mid H_0\}$$

$\Rightarrow$ choose λ, q_1 s.t. any $0 < P_e(H_0) < 1$ can be attained.

for: $P_e(H_0) = 1$: $0 \le \lambda \le a$, q_1 is arbitrary

since $Pr\{\Lambda_{1,0}(R) = \lambda \mid H_0\} = 0$

for: $p_{e0} < P_e(H_0) < 1$: $\lambda = a$, get q_1 from $P_e(H_0) = q_1 \cdot (1-p_{e0}) + p_{e0}$

$$\Rightarrow q_1 = \frac{P_e(H_0) - p_{e0}}{1 - p_{e0}}$$

for: $P_e(H_0) = p_{e0}$: $a < \lambda < b$, $q_1 = 0$

for: $0 < P_e(H_0) < p_{e0}$: $\lambda = b$, $q_1 \cdot p_{e0} = P_e(H_0)$

$$\Rightarrow q_1 = \frac{P_e(H_0)}{p_{e0}}$$

for: $P_e(H_0) = 0$: $\lambda > b$, $q_1 = 0$

<u>Calculation of $P_D = P_c(H_1)$</u>

$$P_{\Lambda_{1,0}}(a \mid H_1) = P_{R|H_1}(0 \mid H_1) = p_{e1}$$

$$P_{\Lambda_{1,0}}(b \mid H_1) = 1 - p_{e1}$$

$$P_c(H_1) = Pr(\Lambda_{1,0} \ge \lambda \mid H_1) \cdot q_1 + Pr(\Lambda_{1,0} > \lambda \mid H_0)\,(1 - q_1)$$

$$= Pr(\Lambda_{1,0} > \lambda \mid H_1) + Pr(\Lambda_{1,0} = \lambda \mid H_1) \cdot q_1$$

<u>Find ROC</u>:

$0 \le \lambda < a$: $P_c(H_1) = 1$

$\lambda = a$: $P_c(H_1) = (1 - p_{e1}) + p_{e1} q_1$

$$= 1 - p_{e1} + p_{e1} \frac{P_e(H_0) - p_{e0}}{1 - p_{e0}} \qquad \text{(linear)}$$

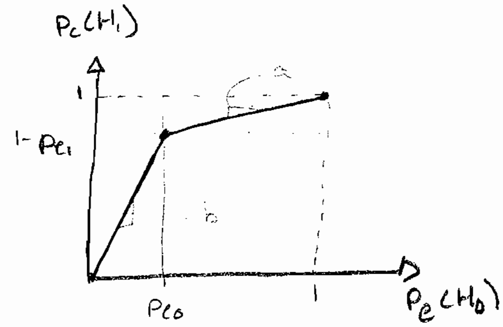
$a < \lambda < b$: $P_c(H_1) = 1 - p_{e1}$

$\lambda = b$: $P_c(H_1) = (1 - p_{e1}) q_1 = (1 - p_{e1}) \dfrac{P_e(H_0)}{p_{e0}}$

his mistake?

$\lambda > b$: $P_e(H_0) = 0$

ROC is:



<u>Notes on ROC:</u>

1) at any point, slope = threshold needed to achieve the point

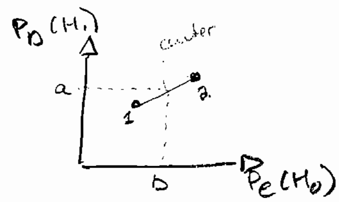
2) concave

$\Rightarrow$ will be proven to be <u>general</u> properties of <u>ROCs</u>

$\boxed{\text{Properties of ROC curves associated with LR test}}$

<u>Theorem 3.2</u>

The ROC curve of a likelihood ratio is "convex down".



Take two points ("tests")

$[P_{c1}(H_1), P_{e1}(H_0)]$, $[P_{c2}(H_1), P_{e2}(H_0)]$

let $a = \dfrac{P_{c1}(H_1) + P_{c2}(H_1)}{2}$

check in book.

$b = \dfrac{P_{e1}(H_0) + P_{e2}(H_0)}{2}$

Construct test 3 where test 1 and test 2 are used with equal probability: $[a, b] = [P_{c3}(H_1), P_{e3}(H_0)]$

From theorem 3.1, know that the LR test with $P_e(H_0) = b$ must have $P_D(H_1) \geq a$ (i.e. $\geq$ any test)

$\rightarrow$ convex down. $\square$

Theorem 3.3

The slope of the ROC curve at any differentiable point is equal to the threshold required to achieve this point.

i.e $\dfrac{d\, P_c(H_1)}{d\, P_e(H_0)} = \lambda$

Proof

1) Assume LR is continuous. ($\rightarrow$ simple decision rule), thus can be characterized by a PDF under H_1, H_0 ; $f_{\Lambda_{10}|H}(\mu, H_m)$, $m=0,$

we have:

$$P_e(H_0) = \int_{\lambda}^{\infty} f_{\Lambda_{10}|H}(\mu|H_0)\, d\mu$$

$$\rightarrow \quad \frac{d\, P_e(H_0)}{d\lambda} = -f_{\Lambda_{10}|H}(\lambda|H_0)$$

Let $Z_1 : \{\underline{r} \text{ s.t. } \Lambda_{10}(\underline{r}) > \lambda\}$ use change of measure formula.

justify more fully.

Then $P_c(H_1) = \underset{\underline{R}}{E}\{1_{Z_1}(\underline{R})|H_1\} = \underset{\underline{R}}{E}[1_{Z_1}(\underline{R})\Lambda_{10}(\underline{R})|H_0]$

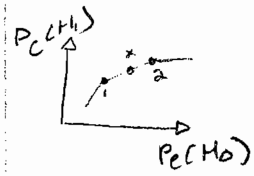
$$= \underset{\Lambda_{10}}{E}[1_{(\lambda,\infty)}(\Lambda_{10}) \cdot \Lambda_{10}|H_0]$$

$$= \int_{\lambda}^{\infty} \mu\, f_{\Lambda_{10}|H}(\mu|H_0)\, d\mu$$

$$\rightarrow \quad \frac{d\, P_c(H_1)}{d\lambda} = -\lambda\, f_{\Lambda_{10}|H}(\lambda|H_0)$$

$$\Rightarrow \quad \frac{d\, P_c(H_1)}{d\, P_e(H_0)} = \lambda$$

2) Assume LR is discrete.



Let the threshold for achieving point 1 be λ_1

The points of abrupt slope change can be achieved a suitable threshold and $q_1 = 0 \Rightarrow$ a simple decision rule.

Other points (on linear segments) we need $q_1 > 0 \Rightarrow$ randomi[zed] decision rule.

At point 1, $q_1 = 0$ and $P_c(H_1) = Pr(\Lambda_{1,0}(\underline{R}) > \lambda_1 \mid H_1)$

$$P_e(H_0) = Pr(\Lambda_{1,0}(\underline{R}) > \lambda_1 \mid H_0)$$

At point x, $q_1 = q_x > 0$ and $P_c(H_1) = Pr(\Lambda_{1,0}(\underline{R}) > \lambda_1 \mid H_1) + q_x Pr(\Lambda_{1,0}(\underline{R}) = \lambda_1 \mid H_1)$

$$P_e(H_0) = Pr(\Lambda_{1,0}(\underline{R}) > \lambda_1 \mid H_0) + q_x Pr(\Lambda_{1,0}(\underline{R}) = \lambda_1 \mid H_0)$$

Slope at point x will be given by

$$m = \frac{P_{cx}(H_1) - P_{c1}(H_1)}{P_{ex}(H_0) - P_{e1}(H_0)}$$

$$= \frac{Pr\{\Lambda_{1,0}(\underline{R}) = \lambda_1 \mid H_1\}}{Pr\{\Lambda_{1,0}(\underline{R}) = \lambda_1 \mid H_0\}} = \frac{Pr[\underline{R} \in Z_{1,0} \mid H_1]}{Pr[\underline{R} \in Z_{1,0} \mid H_0]}$$

$$= \frac{E[1_{Z_{1,0}}(\underline{R}) \mid H_1]}{E[1_{Z_{1,0}}(\underline{R}) \mid H_0]}$$

change of measure formula

$$= \frac{E[\Lambda_{1,0}(\underline{R})\, 1_{Z_{1,0}}(\underline{R}) \mid H_0]}{E[1_{Z_{1,0}}(\underline{R}) \mid H_0]}$$

$$= \frac{\lambda_1 E[1_{Z_{1,0}}(\underline{R}) \mid H_0]}{E[1_{Z_{1,0}}(\underline{R}) \mid H_0]}$$

$$= \lambda_1$$

Note

$1_{Z_{1,0}}(\underline{r})[\Lambda_{1,0}(\underline{r}) - \lambda_1] = 0$

if $\underline{r} \notin Z_{1,0}$, $1_{Z_{1,0}}(\underline{r}) = 0$

if $\underline{r} \in Z_{1,0}$, $\Lambda_{1,0}(\underline{r}) = \lambda_1$

$\Rightarrow E(1_{Z_{1,0}}(\underline{r})[\Lambda_{1,0}(\underline{r}) - \lambda_1]) =$

$\square$

3.4 BAYESIAN FRAMEWORK

Given M Hypotheses $H_0, H_1, \ldots H_M$
generated by a random experiment, occurring with probability P_m (the a priori probabilities) i.e. $P_r(H_m) = P_m$ defined for all m (and known).

Suppose we have a decision rule for the above problem.
We assign a non-negative weight (cost, $0<c<1$) to the events

$$C_{in} = \text{"}H_n \text{ is true and the decision is } H_i\text{"}$$

Which has probability

$$P(H_n \text{ true}, H_i \text{ decided}) = P(H_i | H_n) \cdot P_n$$

The average cost is:

$$\bar{C} = \sum_{i=0}^{M-1} \sum_{n=0}^{M-1} C_{in} P(H_i/H_n) P_n$$

★ The Bayes criterion minimizes $\bar{C}$.

Note: randomized decision rule never used (no improvement gained over simple decision rule)

Using Z_1 to Z_{M-1}

$$\bar{C} = \sum_{i=0}^{M-1} \sum_{m=0}^{M-1} \int_{Z_i} dF_{\underline{R}|H}(\underline{r}|H_m) P_m$$

cost of deciding H_i given the observation.

non-negative

$$= \begin{cases} \sum_{i=0}^{M-1} \int_{Z_i} c_i^c(\underline{r})\, d\underline{r} & , \quad c_i^c(\underline{r}) = \sum_{m=1}^{M-1} C_{im} f_{R|H}(\underline{r}|H_m) P_m \\ \sum_{i=0}^{M-1} \sum_{\underline{r} \in Z_i} c_i^D(\underline{r}) & , \quad c_i^D(\underline{r}) = \sum_{m=1}^{M-1} C_{im} P_{R/H}(\underline{r}|H_m) P_m \end{cases}$$

his mistake.

→ To minimize $\bar{c}$ over the set of all possible partitions, place $\underline{r}$ in region Z_i iff $C_i^c(\underline{r}) = \min\{C_0^c(\underline{r}), \dots C_{M-1}^c(\underline{r})\}$

or $C_i^D(\underline{r}) = \min\{C_0^D(\underline{r}), \dots C_{M-1}^c(\underline{r})\}$

→ This is the optimal Bayes decision rule

<u>Interpretation using Bayes Rule</u>

<u>Continuous</u>

$$\frac{f_{\underline{R}|H}(\underline{r}|H_m)\,p_m}{f_{\underline{R}}(\underline{r})} = P_{H|\underline{R}}(H_m|\underline{r})$$

a priori

<u>Discrete</u>

$$\frac{P_{\underline{R}|H}(\underline{r}|H_m)\,p_m}{P_{\underline{R}}(\underline{r})} = P_{H|\underline{R}}(H_m|\underline{r})$$

a posteriori probabilities H_m given $\underline{r}$.

a priori → [Bayes rule] → a posteriori
measurement →

<u>a posteri costs in terms of a posteriori probs.</u>

$$C_i^{(c,D)}(\underline{r}) = \alpha(\underline{r}) \sum_{m=0}^{M-1} C_{im}\, P_{H|\underline{R}}(H_m|\underline{r})$$

$$\alpha(\underline{r}) = \begin{cases} f_{\underline{R}}(\underline{r}), & \text{continuous} \\ P_{\underline{R}}(\underline{r}), & \text{discrete.} \end{cases}$$

→ Bayes decision rule:

$$\underset{i}{\operatorname{argmin}} \sum_{m=0}^{M-1} c_{im}\, P_{H|\underline{R}}(H_m|\underline{r})$$

For the 2-Hypotheses case : H_0, H_1

Continuous case:

$$\mathcal{C}_0^c(\underline{r}) = C_{00}\, f_{\underline{R}|H}(\underline{r}|H_0)\, p_0 + C_{01}\, f_{\underline{R}|H}(\underline{r}|H_1)\, p_1$$

$$\mathcal{C}_1^c(\underline{r}) = C_{10}\, f_{\underline{R}|H}(\underline{r}|H_0)\, p_0 + C_{11}\, f_{\underline{R}|H}(\underline{r}|H_1)\, p_1$$

Bayes decision rule (through minimization):

$$\mathcal{C}_0^c(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \mathcal{C}_1^c(\underline{r})$$

Equivalent Form:

$$(C_{01} - C_{11})\, f_{\underline{R}|H}(\underline{r}|H_1)\, p_1 \underset{H_0}{\overset{H_1}{\gtrless}} (C_{10} - C_{00})\, f_{\underline{R}|H}(\underline{r}|H_0)\, p_0$$

→ assume $C_{01} > C_{11}$, $C_{10} > C_{00}$

$$\boxed{\underbrace{\frac{f_{\underline{R}|H}(\underline{r}|H_1)}{f_{\underline{R}|H}(\underline{r}|H_0)}}_{\substack{\text{likelihood ratio} \\ \Lambda_{1,0}(\underline{r})}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{(C_{10} - C_{00})\, p_0}{(C_{01} - C_{11})\, p_1}}$$

In terms of a-posteriori probabilities:

$$\boxed{\underbrace{\frac{P_{H|\underline{R}}(H_1|\underline{r})}{P_{H|\underline{R}}(H_0|\underline{r})}}_{\substack{\text{likelihood ratio of} \\ \text{a posteriori probs} \\ \Omega_{1,0}(\underline{r})}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{(C_{10} - C_{00})}{(C_{01} - C_{11})}}$$

→ turbo coding notation

Minimization of average probability of error

Choose: $C_{in} = \begin{cases} 1, & i \neq n \\ 0, & i = n \end{cases}$

leads to: $\bar{C} = \sum_{i=0}^{M-1} \sum_{\substack{m=0 \\ m \neq i}}^{M-1} P(H_i|H_m) p_m$

$= \sum_{\substack{m=0 \\ m \neq i}}^{M-1} P_e(H_m) p_m$

$= P_e$, the "average probability of error".

where $P_e(H_m) = \sum_{\substack{i=0 \\ i \neq m}}^{M-1} P(H_i|H_m)$, the "probability of error associated with H_m."

// Insert lecture 9 here.

Consider the set of all cost vectors $\underline{V}_c$ s.t. for a given $\underline{P}$ and $b \geq 0$, $\underline{V}_c$ satisfies $\underline{P}^T \underline{V}_c = b$

→ $\underline{V}_c$ is a plane (or hyperplane of dim. $M-1$), normal $\frac{\underline{P}}{\|\underline{P}\|}$
→ all decision rules associated with those $\underline{V}_c$ are equivalent since $\bar{C} = b$

$\boxed{\text{Def}}$ A hyperplane is said to support the convex set S_v at the boundary point $\underline{V}_c^0$, if $\underline{V}_c^0$ is a boundary point for S_v, and the following holds:

support hyperplane

1- Hyperplane passes through $\underline{V}_c^0$
2- S_v is on one side of the hyperplane.

Property of convex sets

Any convex set has a supporting hyperplane at any boundary point.

Note supporting plane ≠ tangential plane ?

many supporting planes but no tangential plane.

- Optimal Bayes decision rules correspond to points on the boundary of S_v.
- Given a point on the boundary corresponding to an optimal Bayes rule, we can find its supporting hyperplane, which has normal = associated vector of a-prior probabilities.

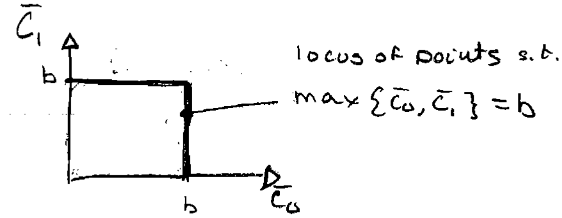
The Minimax strategy

What if the a prior probabilities P_i are unknown?

→ Find the decision rule which minimizes the maximum of the a posterior cost.

$$\min \max \{\bar{C}_0, \dots \bar{C}_{M-1}\}$$

When $M=2$ (two hypothesis)



→ Can be extended to higher dimensions.

- The minmax decision $\underline{V}_c^{minmax}$ is characterized by

$$\underline{V}_c^{minmax} = \bar{C}_0 [1 \ 1 \dots 1]^T$$

→ called equalization of conditional costs.

- The a prior prob. vector $\underline{P}$ associated with $\underline{V}_c^{minmax}$ is called the most unfavorable a-prior, and must satisfy $P_0 > 0, \dots P_{M-1} > 0$.

<u>Note</u>: sometimes applying min max gives some $P_i < 0$, in which case we say that the problem does not have a minimax decision rule.

More on most-unfavorable a-prior

Let $\bar{C}_{min}(\underline{P})$ denote the minimum average cost for a-prior vector $\underline{P}$.

Then

$$\bar{C}_{min}(\underline{P}) = \underline{P}^T \cdot \underline{V}_c(\underline{P})$$

*($\bar{V}_c(\underline{P})$ also depends on $\underline{P}$ since to achieve $\bar{C}_{min}$ it must correspond to an optimal Bayes test for a prior $\underline{P}$)

Theorem 3.5.2

$\bar{C}_{min}(\underline{P})$ is convex down $\cap$ function of $\underline{P}$.

Proof:

Consider two a-prior probability vectors $\underline{P}^1, \underline{P}^2$ and the corresponding minimum average costs.

$$C_{min}(\underline{P}^1) = \underline{P}^{1T} V_c(\underline{P}^1)$$

$$C_{min}(\underline{P}^2) = \underline{P}^{2T} V_c(\underline{P}^2)$$

$$C_{min}\left(\tfrac{1}{2}(\underline{P}^1+\underline{P}^2)\right) = \tfrac{1}{2}(\underline{P}^1+\underline{P}^2)^T \, V_c\left(\tfrac{1}{2}(\underline{P}^1+\underline{P}^2)\right)$$

$$= \tfrac{1}{2}\underline{P}^{1T} \underline{V}_c\left(\tfrac{1}{2}(\underline{P}^1+\underline{P}^2)\right)$$

$$+ \tfrac{1}{2}\underline{P}^{2T} \underline{V}_c\left(\tfrac{1}{2}(\underline{P}^1+\underline{P}^2)\right)$$

$$\geq \tfrac{1}{2}\left[C_{min}(\underline{P}^1) + C_{min}(\underline{P}^2)\right]$$

Corollary

$\bar{C}_{min}(\underline{P})$ has only one maximum

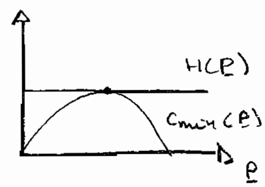
(follows from convexity down)

Assume now that $C_{min}(\underline{P})$ achieves its maximum at $\underline{P}^o$ with non-zero components.

Consider the support hyperplane of $C_{min}(\underline{P})$ at $\underline{P}^o$

? $\rightarrow$ $$H(\underline{P}) = \underline{P}^t \underline{V}_c(\underline{P}^o)$$

Since it is tangential to $C_{min}(\underline{P})$ at its maximum, then $H(\underline{P})$ cannot vary with $\underline{P}$



<u>Claim</u>: $V_c(\underline{P})$ has equal components.

<u>Outline of proof</u>:

- Choose $0 < \varepsilon < \frac{1}{M-1}$
- Define the apriori vector set

$$\underline{P}^{1T} = [\,1-(M-1)\varepsilon,\ \varepsilon,\ \varepsilon \ldots \varepsilon\,]$$
$$\underline{P}^{2T} = [\,\varepsilon,\ 1-(M-1)\varepsilon,\ \varepsilon,\ \varepsilon \ldots \varepsilon\,]$$
$$\vdots$$
$$\underline{P}^{MT} = [\,\varepsilon,\ \varepsilon,\ \cdots,\ \varepsilon,\ 1-(M-1)\varepsilon\,]$$

$\Rightarrow$ Then, $H(\underline{P}^1) = H(\underline{P}^2) = \cdots = H(\underline{P}^M) = H_{const}$.

In matrix form

$$\begin{bmatrix} \underline{P}^{1T} \\ \underline{P}^{2T} \\ \vdots \\ \underline{P}^{MT} \end{bmatrix} \cdot V_c(\underline{P}^o) = H_{const} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

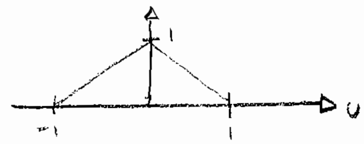
$\Rightarrow$ $V_c(\underline{P}^o)$ has equal components.

// Feb. 11

Minimax example.

$\boxed{\text{Ex}}\ \boxed{3.5.1}$

Let $f(w)$ be a pdf defined by:



And consider the three-hypothesis testing problem:

$H_0: \ x = w$

$H_1: \ x = w + s \qquad 1 < s < 2$ is known constant

$H_2: \ x = w - s \qquad w$ is noise with pdf $f(w)$

Find the minimax decision rule minimizing the average prob. of error (so set $C_{ij} = 1, \ i \neq j$; $C_{ij} = 0, \ i = j$)

$f_{x|H_0}(x|H_0) = f(x)$

$f_{x|H_1}(x|H_1) = f(x-s)$

$f_{x|H_2}(x|H_2) = f(x+s)$

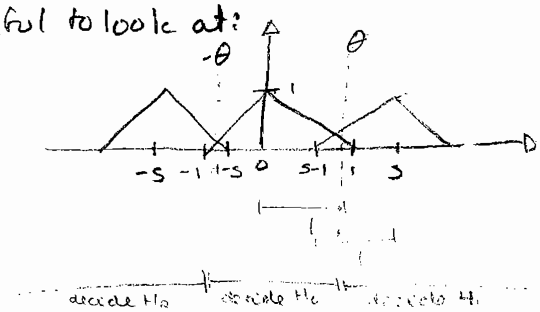
To find minimax decision rule:
Set average conditional costs equal

for $m = 0,1,2 \quad \bar{C}_m = \sum_{\substack{i=0 \\ i \neq m}}^{2} P(H_i|H_m)$

$$P(H_i|H_m) = \int_{Z_i} f_{x|H_m}(x|H_m)\, dx$$

$\rightarrow$ find a partition Z_i satisfying $\bar{C}_0 = \bar{C}_1 = \bar{C}_2$

Useful to look at:



$\rightarrow$ choose θ to satisfy $\bar{C}_0 = \bar{C}_1 = \bar{C}_2$

(it is easily guessed that $s-1<\Theta<1$
which leads us to:

$$\bar{C}_0 = P(H_1|H_0) + P(H_2|H_0)$$

$$= \int_{\Theta}^{1} x+1-s \, dx + \int_{-1}^{-\Theta} x+1 \, dx$$

$$= \left[\tfrac{1}{2}x^2 + (1-s)x \right]_{\Theta}^{1} + \left[\tfrac{1}{2}x^2 + x \right]_{-1}^{-\Theta}$$

$$= \left[\cancel{\tfrac{1}{2}} + 1 - s - \tfrac{1}{2}\Theta^2 + (s-1)\Theta \right] + \left[\tfrac{1}{2}\Theta^2 - \Theta - \tfrac{1}{2} + 1 \right]$$

$$= (s-2)\Theta + 2 - s$$

$$\bar{C}_1 = P(H_0|H_1) + P(H_2|H_1)$$

$$= \int_{s-1}^{\Theta} -x+1 \, dx + 0$$

$$= \left[-\tfrac{1}{2}x^2 + x \right]_{s-1}^{\Theta}$$

$$= -\tfrac{1}{2}\Theta^2 + \Theta + \tfrac{1}{2}(s-1)^2 + 1 - s$$

$$\bar{C}_2 = P(H_0|H_2) + P(H_1|H_2)$$

$$= \int_{-\Theta}^{1-s} x+1 \, dx + 0$$

$$= \left[\tfrac{1}{2}x^2 + x \right]_{-\Theta}^{1-s}$$

$$= \tfrac{1}{2}(1-s)^2 + 1 - s - \tfrac{1}{2}\Theta^2 + \Theta$$

same as (makes sense!)

His final answer

$P(H_1|H_0) = P(H_2|H_0) = \tfrac{1}{2}(1-\Theta$

$P(H_0|H_1) = P(H_0|H_2) = \tfrac{1}{2}(1+\Theta-s$

$\bar{C}_0 = (1-\Theta)^2$

$\bar{C}_1 = \bar{C}_2 = \tfrac{1}{2}(1+\Theta-s)^2$

Cost equalization

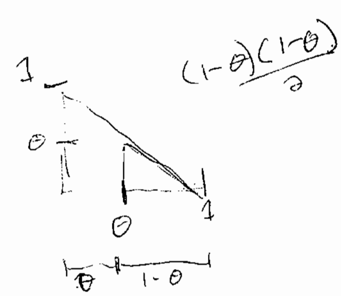
$2(1-\Theta)^2 = (1+\Theta-s)^2$

Two solutions

$\sqrt{2}(1-\Theta) = \pm(1+\Theta-s)$

$\Theta_1 = \frac{\sqrt{2}-1}{\sqrt{2}+1} + \frac{s}{\sqrt{2}+1}$

is the valid answer.



We have $\bar{C}_1 = \bar{C}_2$ always,

So solve $\bar{C}_0 = \bar{C}_1$ for Θ.

$$(s-2)\Theta + 2 - s = -\tfrac{1}{2}\Theta^2 + \Theta + \tfrac{1}{2}(s-1)^2 + 1 - s$$

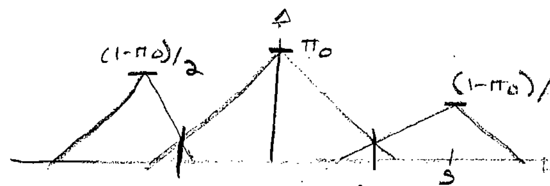
$$\tfrac{1}{2}\Theta^2 + (s-3)\Theta + \left[1 + \tfrac{1}{2}(s-1)^2 \right] = 0$$

$$-(s-3) \pm \sqrt{(s-3)^2 - 2\left[1 + \tfrac{1}{2}(s-1)^2\right]}$$

$(\ldots)$)

Now, find the most unfavorable a priori probabilities.

→ it is the apriori that generates the minimax decision rule.
Based on Bayes.



Θ = Bayes threshold.

→ what weighing (scaling) makes the intersections (which define various Bayes thresholds) correspond to the minimax threshold Θ? (& assume symmetry)

intersection of lines:

$$\pi_0 - \pi_0 \Theta = (\pi_0 - 1)/2 \cdot (S-1) + \frac{(1-\pi_0)}{2}\Theta$$ ← check

$$\pi_0 = \frac{\Theta - S + 1}{2 - S}$$ ← his

$$\Theta = \frac{\sqrt{2}-1}{\sqrt{2}+1} + \frac{S}{\sqrt{2}+1}$$

(...)

$$\boxed{\begin{aligned} \pi_0 &= \frac{1}{\sqrt{2}} \\ \pi_2 = \pi_1 &= 1 - \pi_0 \end{aligned}}$$

$\boxed{\text{Composite Hypothesis testing}}$

Each Hypothesis induces a family of probability laws.

$$H_i : F_{\underline{R}|H,\underline{B}}(\underline{r}|H_m, \underline{b})$$

where $\underline{B}$ is a Q dimensional parameter vector.

<u>Bayesian Case</u>

We consider $\underline{b}$ to be a realization of a random vector $\underline{\mathbf{B}}$, which under hypothesis H_i has cdf $F_{\underline{B}|H}(\underline{b}|H_i)$

<u>Define</u> $\boxed{0 \le C_{im}(b) \le 1}$, cost of event "$H_m$ is true & decision is H_i given $\underline{B} = \underline{b}$"

$$\boxed{\bar{C}_m = \sum_{i=0}^{M-1} C_{im}(\underline{b}) P(H_i | H_m, \underline{b})}$$ "conditional cost given H_m is true and $\underline{B} = \underline{b}$"

$P(H_i|H_m, \underline{b})$ "prob. decide H_i given H_m is true and $\underline{B} = \underline{b}$.

<u>Note</u>:

$$P(H_i|H_m,\underline{b}) = \int_{Z_i} dF_{\underline{R}|H,\underline{B}}(\underline{r}|H_m,\underline{b})$$

$$= \begin{cases} \int_{Z_i} f_{\underline{R}|H,\underline{B}}(\underline{r}|H_m,\underline{b})\, d\underline{r} \\ \sum_{r \in Z_i} P_{\underline{R}|H,\underline{B}}(\underline{r}|H_m,\underline{b}) \end{cases}$$

Derivation of average cost. (averaged also over $\underline{B}$)

Conditional average cost, given H_m is true:

$$\bar{C}_m = E[\bar{C}_m(\underline{B}) \mid H_m]$$
$$= \sum_{i=0}^{M-1} E[C_{im}(\underline{B}) P(H_i \mid H_m, \underline{B}) \mid H_m]$$

Average cost:

$$\bar{C} = \sum_{m=0}^{M-1} \bar{C}_m P_m$$
$$= \sum_{m=0}^{M-1} \sum_{i=0}^{M-1} E[C_{im}(\underline{B}) P(H_i \mid H_m, \underline{B}) \mid H_m] P_m$$

Need to find the decision rule which minimizes $\bar{C}$

Continuous case:

Note: this is the only new thing in composite hypothesis testing.

$$\bar{C} = \sum_{m=0}^{M-1} \sum_{i=0}^{M-1} E[C_{im}(\underline{B}) \int_{Z_i} f_{\underline{R}|H,\underline{B}}(\underline{r} \mid H_m, \underline{B}) d\underline{r} \mid H_m] P_m$$

$$= \sum_{i=0}^{M-1} \int_{Z_i} C_i^c(\underline{r}) d\underline{r}$$

where the costs when we decide H_i given the observation $\underline{r}$ are given by

$$C_i^c(\underline{r}) = \sum_{m=0}^{M-1} E[C_{im}(\underline{B}) f_{\underline{R}|H,\underline{B}}(\underline{r} \mid H_m, \underline{B}) \mid H_m] P_m$$

which is a non-negative function of $\underline{r}$.

Discrete case:

$$\bar{C} = \sum_{m=0}^{M-1} \sum_{i=0}^{M-1} C_i^D(\underline{r})$$

$$C_i^D(\underline{r}) = \sum_{m=0}^{M-1} E[C_{im}(\underline{B}) P_{\underline{R}|H,\underline{B}}(\underline{r} \mid H_m, \underline{B}) \mid H_m] P_m.$$

Since non-negative, we have the optimal decision rule:

Decide H_i if $C_i^c(\underline{r}) = \min(C_0^c(\underline{r}), \ldots, C_{M-1}^c(\underline{r}))$
$C_i^D(\underline{r}) = \min(C_0^D(\underline{r}), \ldots, C_{M-1}^D(\underline{r}))$ } hard to calculate.

→ like Bayesian.

Interpretation using Bayes rule

Continuous case

$$f_{\underline{R}|H,\underline{B}}(\underline{r}|H_m,\underline{b})\cdot P_m = f_{\underline{R}|\underline{B}}(\underline{r}|\underline{b})\,P_{H|\underline{R}}(H_m|\underline{r},\underline{b})$$

So write:

$$C_i^{\,c}(\underline{r}) = \sum_{m=0}^{M-1} E\left[C_{im}(\underline{B})\, f_{\underline{R}|\underline{B}}(\underline{r}|\underline{B})\,P_{H|\underline{R},\underline{B}}(H_m|\underline{r},\underline{b})\right]$$

Discrete case:

$$C_i^{\,D}(\underline{r}) = \sum_{m=0}^{M-1} E\left[C_{im}(\underline{B})\, P_{\underline{R}|\underline{B}}(\underline{r}|\underline{B})\,P_{H|\underline{R},\underline{B}}(H_m|\underline{r},\underline{b})\right]$$

→ compare with simple hypothesis testing.

→ (because of dependence of C_{im} on $\underline{B}$), cannot have as before completely a-posteriori probabilities.

Special Case: choose the weights $C_{im}(\underline{B})$ not to depend on $\underline{B}$, i.e. $C_{im}(\underline{b}) = C_{im}$

→ becomes simple hypothesis testing problem.

Because:

$$\bar{C} = \sum_{m=0}^{M-1}\sum_{i=0}^{M-1} C_{im}\, E\left[P(H_i|H_m,\underline{B})\,|\,H_m\right] P_m$$

i.e. can take C_{im} out of expectation.

$$= \sum_m \sum_i C_{im}\, P(H_i|H_m)$$

$$= \begin{cases} \sum_m \sum_i C_{im} \int_{Z_i} f_{\underline{R}|H}(\underline{r}|H_m)\,d\underline{r} \\ \sum_m \sum_i C_{im} \sum_{\underline{r}\in Z_i} P_{\underline{R}|H}(\underline{r}|H_m) \end{cases}$$

→ simple hypothesis testing with

$$f_{\underline{R}|H}(\underline{r}|H_m) = \int f_{\underline{R}|H,\underline{B}}(\underline{r}|H_m,\underline{b})\,dF_{\underline{B}|H}(\underline{b}|H_m)$$

$$P_{\underline{R}|H}(\underline{r}|H_m) = \int P_{\underline{R}|H,\underline{B}}(\underline{r}|H_m,\underline{b})\,dF_{\underline{B}|H}(\underline{b}|H_m)$$

→ under each hypothesis, the parameter $\underline{B}$ is averaged out from $F(\underline{r}|H_m, \underline{b})$ using pdf induced by this hypothesis $F_{\underline{B}|H_m}(\underline{b}|H_m)$, resulting in a simple hypotheses testing

To Summarize

In the Bayesian Framework, when the weights c_{im} do not depend on $\underline{B}$, we can transform the composite hypothesis testing problem into a simple hypothesis problem.

Composite:

$$H_m : \left\{ \begin{array}{l} f_{\underline{R}|H,\underline{B}}(\underline{r}|H_m, \underline{b}) \\ P_{\underline{R}|H,\underline{B}}(\underline{r}|H_m, \underline{b}) \end{array} \right\}, \quad F_{\underline{B}|H}(\underline{b}|H_m)$$

becomes:

Simple:

$$H_m : \left\{ \begin{array}{l} \int f_{\underline{R}|H,\underline{B}}(\underline{r}|H_m, \underline{b}) \, dF_{\underline{B}|H}(\underline{b}|H_m) \\ \int P_{\underline{R}|H,\underline{B}}(\underline{r}|H_m, \underline{b}) \, dF_{\underline{B}|H}(\underline{b}|H_m) \end{array} \right\}$$

Neyman - Pearson composite hypothesis testing (2 hypothesis)

→ $\underline{B}$ is an unknown parameter.

$$H_1: F_{\underline{R}|H,\underline{B}}(\underline{r}|H_1,\underline{b}_1), \quad b_1 \in \beta_1 \subseteq \mathbb{R}^Q$$
$$H_0: F_{\underline{R}|H,\underline{B}}(\underline{r}|H_0,\underline{b}_0), \quad b_0 \in \beta_0 \subseteq \mathbb{R}^Q$$

→ i.e. known what the possible values of $\underline{b}_i$ are, but no idea of its probability distribution function.

Def. A uniformly Most powerful (UMP) test is a test such that for all possible values of $\underline{b}_1$ and $\underline{b}_0$, it maximizes the probability of detection under the constraint that the probability of false alarm $P_F \le \alpha$.

Theorem 3.6.1

A UMP of level α exists iff a Neyman-Pearson optimal test of level α can be constructed such that it does not depend on $\underline{b}_0, \underline{b}_1$ for all $\underline{b}_0 \in \beta_0$, $\underline{b}_1 \in \beta_1$

Cases where UMP may exist (note: ordering of H_1, H_0 is important)

① $H_1: f_{\underline{R}|H,\underline{B}}(\underline{r}|H_1,\underline{b}_1)$, $\underline{b} \in \beta_1 \subseteq \mathbb{R}^Q$ omposite)

$H_0: f_{\underline{R}|H}(\underline{r}|H_0)$ (simple)

In that case:

$$\Lambda_{1,0}(\underline{r},\underline{b}_1) = \frac{f_{\underline{R}|H,\underline{B}}(\underline{r}|H_1,\underline{b}_1)}{f_{\underline{R}|H}(\underline{r}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

$$\alpha = P_{FA} = \int_{Z_1} f_{\underline{R}|H}(\underline{r}|H_0)\, d\underline{r} \quad \rightarrow \text{does not depend on } \underline{b}_1$$

→ The decision regions & therefore the test do not depend on $\underline{b}_1$, so <u>we have UMP test</u>

$\boxed{\text{Example}}$ 3.6.1

<u>Define</u> $\quad \underline{R} \sim f(r,\theta) = \begin{cases} \theta e^{-r\theta}, & r \geq 0 \\ 0, & r < 0 \end{cases}$

With parameter $\theta > 0$ (unknown).

Formulate the hypothesis testing problem;

$H_1: \; f_{R|H}(r|H_1) = f(r,\theta), \; \theta > \theta_0$ → composite

$H_0: \; f_{R|H}(r|H_0) = f(r,\theta_0)$ → single.

And $\theta_0 > 0$ is given.

Neyman-Pearson test;

$$\Lambda_{1,0}(\underline{r},\theta) = \frac{f(r,\theta)}{f(r,\theta_0)}$$

$$= \frac{\theta e^{-r\theta}}{\theta_0 e^{-r\theta_0}}$$

$$= \theta/\theta_0 \; e^{r(\theta_0-\theta)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

$$\equiv -(\theta-\theta_0) r \underset{H_0}{\overset{H_1}{\gtrless}} \ln\left(\lambda \frac{\theta_0}{\theta}\right)$$

$$\underbrace{\boxed{r \underset{H_1}{\overset{H_0}{\gtrless}} \frac{1}{\theta-\theta_0} \ln\left(\frac{\theta}{\lambda\theta_0}\right)}}_{\lambda'}$$

$$\alpha = \int_{Z_1} f(\underline{r}|H_0)\, d\underline{r}$$

$$= \int_0^{\lambda'} \theta_0 e^{-r\theta_0}\, d\underline{r}$$

$$= 1 - e^{-\theta_0 \lambda'}$$

$\rightarrow \boxed{\lambda' = \frac{1}{\theta_0}\ln(1-\alpha)}$ $\rightarrow$ test does not depend on θ_0

$\rightarrow$ UMP test exists.

$\boxed{\text{Example}}$ 3.6.2

Same as 3.6.1, except:

H_1: $f(\underline{r}|H_1,\theta) = f(r,\theta)$ $\quad \theta > 0$, $\theta \neq \theta_0$ $\quad$ (do not know $\theta > \theta_0$)

H_0: $f(r|H_0) = f(r,\theta_0)$

$$\Lambda_{g0}(\underline{r},\theta) = \begin{cases} r \underset{H_1}{\overset{H_0}{\gtrless}} \frac{1}{\theta - \theta_0} \ln\left(\frac{\theta}{\lambda\theta_0}\right), & \text{if } \theta > \theta_0 \\ r \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{\theta - \theta_0} \ln\left(\frac{\theta}{\lambda\theta_0}\right), & \text{if } \theta < \theta_0 \end{cases}$$

<u>case $\theta > \theta_0$</u>

$$\alpha = \int_0^{\lambda'} \theta_0 e^{-r\theta_0} d\underline{r} \rightarrow \lambda' = \frac{1}{\theta_0}\ln(1-\alpha)$$

<u>case $\theta < \theta_0$</u>

$$\alpha = \int_{\lambda'}^{\infty} \theta_0 e^{-r\theta_0} d\underline{r} \rightarrow \lambda' = -\frac{1}{\theta_0}\ln(\alpha)$$

$$= e^{-\lambda'\theta_0}$$

$\rightarrow$ threshold depends on θ.

(actually, on sign of $\theta - \theta_0$)

In cases where UMP test does not exist:

$\boxed{\text{Locally most powerful test}}$

Consider the hypothesis testing problem:

$$H_1 : f_{\underline{R}|H,B}(\underline{r}|H_1, b_1) = f(\underline{r}, b_1) \quad , b_1 > \theta_0$$
$$H_0 : f_{\underline{R}|H}(\underline{r}|H_0) = f(\underline{r}, b_0) \quad , b_0 = \theta_0$$

Consider a decision rule with a certain probability of detection P_D. In general, P_D will depend on b_1.

Using Taylor's expansion around $P_D(\theta_0)$

$$P_D(b_1) \simeq P_D(\theta_0) + (b_1 - \theta_0) \left. \frac{\partial P_D(b)}{\partial b} \right|_{b=\theta_0}$$

$\alpha \cdot \cos)$

$$P_D(\theta_0) = \int_{Z_1} f_{R|H_1,B}(\underline{r}|H_1, \theta_0)\, d\underline{r}$$
$$= \int_{Z_1} f(\underline{r}, \theta_0)\, d\underline{r}$$
$$= \int_{Z_1} f_{\underline{R}|H,B}(r, H_0, b_0)$$
$$= P_{FA}$$

$$\text{So: } \quad P_D(b_1) \simeq P_{FA} + \overbrace{(b_1 - \theta_0)}^{\text{positive}} \left. \frac{\partial P_D(b)}{\partial b} \right|_{b=\theta_0}$$

Maximize P_D subject to $P_{FA} \le \alpha$

Can maximize $\left. \frac{\partial P_D(b)}{\partial b} \right|_{b=\theta_0}$ to approximately maximize $P_D(b_1)$ for all $b_1 > \theta_0$

Structure of LMP test

Find the partition Z_0, Z_1 that maximizes

$$\left.\frac{\partial P_D}{\partial b}\right|_{b=\Theta_0} = \int_{Z_1} \left.\frac{\partial f(\underline{r}, b)}{\partial b}\right|_{b=\Theta_0} d\underline{r}$$

Under the constraint

$$P_{FA} = \int_{Z_1} f(\underline{r}, \Theta_0)\, d\underline{r} \leq \alpha .$$

→ This can be solved using similar methods as in Neyman-Pearson theorem. (3,1):

$$\boxed{\frac{\left.\frac{\partial f(\underline{r}, b)}{\partial b}\right|_{b=\Theta_0}}{f(\underline{r}, \Theta_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda}$$

with λ set such that:

$$\boxed{\alpha = \int_{Z_1} f(\underline{r}, \Theta_0)\, d\underline{r}}$$

Equivalent form:

$$\boxed{\left.\frac{\partial \ln f(\underline{r}; b)}{\partial b}\right|_{b=\Theta_0} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda}$$

(Could be solved in Bayesian, with some assumptions)

$\boxed{Ex}$ 3.6.3

Consider $\underline{R} = [R_1 \; R_2]^T$

parameter

$$H_1: \; f_{\underline{R}|H,B}(\underline{r}|H_1, \rho) = \frac{1}{2\pi\sqrt{\det C_1}} \exp\left(-\tfrac{1}{2}\underline{r}^T C_1^{-1}\underline{r}\right)$$

$$H_0: \; f_{\underline{R}|H,B}(\underline{r}|H_0, \rho) = \frac{1}{2\pi\sqrt{\det C_0}} \exp\left(-\tfrac{1}{2}\underline{r}^T C_0^{-1}\underline{r}\right)$$

Where $\quad C_1 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad C_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad 0<\rho<1$

i.e. are R_1 & R_2 correlated?

① Try UMP,

$$\Lambda_{1,0}(\underline{r}) = \sqrt{\frac{\det C_0}{\det C_1}} \cdot \exp\left(-\tfrac{1}{2}\underline{r}^T[C_1^{-1} - C_0^{-1}]\underline{r}\right)$$

$$= \frac{1}{\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{\rho}{2(1-\rho^2)}\underline{r}^T \begin{bmatrix} \rho & -1 \\ -1 & \rho \end{bmatrix}\underline{r}\right)$$

$$\ln \Lambda_{1,0}(\underline{r}) = -\tfrac{1}{2}\ln(1-\rho^2) - \frac{\rho}{2(1-\rho^2)}\left[\rho r_1^2 + \rho r_2^2 - 2r_1 r_2\right]$$

$$\ln \Lambda_{1,0}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

$$\frac{-\rho}{2(1-\rho^2)}(\rho r_1^2 + \rho r_2^2 - 2r_1 r_2) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda + \ln(1-\rho^2)$$

$$\rho r_1^2 + \rho r_2^2 - 2r_1 r_2 \underset{H_1}{\overset{H_0}{\gtrless}} -\frac{2(1-\rho)}{\rho}\left(\lambda + \tfrac{1}{2}\ln(1-\rho^2)\right)$$

→ depends on ρ.

→ UMP does not exist.

② Resort to LMP (sub-optimal)

$$\ln p_{\underline{r}|H_1,B}(\underline{r}|H_1,\rho) = -\ln\left(2\pi\sqrt{\det C_1}\right) - \frac{1}{2(1-\rho^2)}\,\underline{r}^T\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}\underline{r}$$

$$\frac{\partial \ln(\underline{r}|H_1,\rho)}{\partial\rho} = -\tfrac{1}{2}\frac{\partial}{\partial\rho}\left[\frac{r_1^2+r_2^2-2\rho r_1 r_2}{1-\rho^2}\right]$$

$$= -\tfrac{1}{2}\,\frac{1-\rho^2-2r_1r_2+2r_1^2\rho+2r_2^2\rho-4\rho r_1 r_2}{(1-\rho^2)^2}$$

$$\left.\frac{\partial \ln(\underline{r}|H_1,\rho)}{\partial\rho}\right|_{\rho=0} = -\tfrac{1}{2}+r_1r_2$$

LMP test is:

$$\boxed{r_1 r_2 \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{\lambda+\tfrac{1}{2}}_{\lambda'}}$$

Set threshold according to probability of false alarm

$$\alpha = \text{Prob}(r_1 r_2 > \lambda' \mid H_0)$$

↳ will need to numerically evaluate a Bessel function.

★ Remember that LMP test is sub-optimal in Neyman-Pearson sense: P_D is not exactly maximized.

Another sub-optimal approach:

Generalized Likelihood approach

Consider the M-ary hypotheses problem:

$$m = 0, 1, \ldots, M-1$$

$$H_m = f_{\underline{R}|H,\underline{B}}(\underline{r} \mid H_m, \underline{b}_m), \quad \underline{b}_m \in \beta_m \subseteq \mathbb{R}^Q$$

let: ..

Continuous case:

let $\hat{\underline{b}}_m \in \beta_m$ be such that

$$f_{\underline{R}|H,\underline{B}}(\underline{r} \mid H_m, \hat{\underline{b}}_m) = \max_{\underline{b} \in \beta_m} f_{\underline{R}|H,\underline{B}}(\underline{r} \mid H_m, \underline{b})$$

Discrete case:

let $\hat{\underline{b}}_m \in \beta_m$ be such that

$$P_{\underline{R}|H,\underline{B}}(\underline{r} \mid H_m, \hat{\underline{b}}_m) = \max_{\underline{b} \in \beta_m} P_{\underline{R}|H,\underline{B}}(\underline{r} \mid H_m, \underline{b})$$

for $m = 0, 1, \ldots M-1$

→ $\hat{\underline{b}}_m$ is a maximum likelihood (ML) estimate of $\underline{b}$ under hypothesis H_m.

Key point

→ replace the original composite hypothesis testing problem by a simple hypothesis testing problem:

$$H_m : F_{\underline{R}|H,\underline{B}}(H_m, \hat{\underline{b}}_m)$$

(4) DISCRETE TIME SIGNAL DETECTION

4.1 Detection of deterministic signals in gaussian noise

Consider M discrete time, known signals:

$$\underline{s}_m = [s_m(0), s_m(1), \ldots, s_m(1-n)]^T, \; m = 0, 1, \ldots M-1$$

Assume that the observed vector is

$$\underline{r} = [r(0), r(1), \ldots, r(1-n)]^T$$

where, under Hypothesis H_m is given by the following:

$$H_m: \underline{r} = \underbrace{\underline{s}_m}_{\text{one of the M known vectors.}} + \underline{n}$$

with $\underline{n}$ as the noise vector, modeled as a sample of a Gaussian random vector $\underline{N}$ with mean zero and known covariance $E[\underline{N}\,\underline{N}^T] = C_N$, $C_N > 0$.

(C_N [illegible] always full rank)

Hence:

$$H_m: f_{\underline{R}|H}(\underline{r}|H_m) = \frac{1}{\sqrt{(2\pi)^{n} \det(C_N)}} \cdot \exp\left[-\tfrac{1}{2}(\underline{r}-\underline{s}_m)^T \cdot C_N^{-1}(\underline{r}-\underline{s}_m)\right]$$

mean = signal corr. to H_m

Case $M=1$

Likelihood ratio:

$$\Lambda_{1,0}(\underline{r}) = \frac{f_{\underline{R}|H}(\underline{r}|H_1)}{f_{\underline{R}|H}(\underline{r}|H_0)} = \frac{\exp[-1/2(\underline{r}-\underline{s}_1)C_N^{-1}(\underline{r}-\underline{s}_1)]}{\exp[-1/2(\underline{r}-\underline{s}_0)C_N^{-1}(\underline{r}-\underline{s}_0)]}$$

$$= \frac{\exp[-1/2\,\underline{r}^TC_N^{-1}\underline{r} + 1/2\,\underline{r}^TC_N^{-1}\underline{s}_1 + 1/2\,\underline{s}_1^TC_N^{-1}\underline{r} - 1/2\,\underline{s}_1^TC_N^{-1}\underline{s}_1]}{\exp[-1/2\,\underline{r}^TC_N^{-1}\underline{r} + 1/2\,\underline{r}^TC_N^{-1}\underline{s}_0 + 1/2\,\underline{s}_0^TC_N^{-1}\underline{r} - 1/2\,\underline{s}_0^TC_N^{-1}\underline{s}_0]}$$

use $C_N = C_N^T$
$C_N^{-1} = C_N^{-1T}$
$r^TC_N^{-1}s_M$ is scalar.
$\to$ sym.

$$= \exp[\,\underline{r}^TC_N^{-1}(\underline{s}_1-\underline{s}_0) + 1/2\,\underline{s}_0^TC_N^{-1}\underline{s}_0 - 1/2\,\underline{s}_1^TC_N^{-1}\underline{s}_1]$$

Likelihood test:

$$\Lambda_{1,0}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

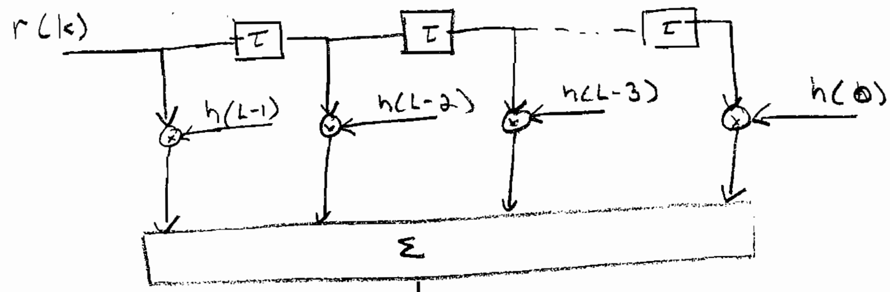
Log likelihood test:

$$\underbrace{\underline{r}^T\overbrace{C_N^{-1}(\underline{s}_1-\underline{s}_0)}^{\underline{h}}}_{\to\text{ linear receiver}} \underset{H_0}{\overset{H_1}{\gtrless}} \log\lambda + 1/2\,\underline{s}_1^TC_N^{-1}\underline{s}_1 - 1/2\,\underline{s}_0^TC_N^{-1}\underline{s}_0 = \lambda'$$

Notes: $\underline{r}^T\underline{h}$ is a linear filtering of the observation by the filter $\underline{h}$

- $\underline{h}$ computed from $C_N\underline{h} = (\underline{s}_1-\underline{s}_0)$
- $\underline{h}$ is a matched filter to $(\underline{s}_1-\underline{s}_0)$
- White noise $\to$ $C_N = I$ $\to$ $\underline{h} = (\underline{s}_1-\underline{s}_0)$ $\}\to$ in Kay.

Implementation



$$\sum_{k=0}^{L-1} r(k)\,h(k) = \underline{r}^T \underline{h}$$

(output valid when all values are loaded)

Performance for binary hypothesis

$$P(H_1|H_0) = Pr\left[\underline{R}^T C_N^{-1}(\underline{s}_1-\underline{s}_0) > \lambda' \,|\, H_0\right]$$

Gaussian, mean $\underline{s}_0^T C_N^{-1}(\underline{s}_1-\underline{s}_0)$

var: $(\underline{s}_1-\underline{s}_0)^T C_N^{-1}(\underline{s}_1-\underline{s}_0)$

$$= Q\left[\frac{\lambda' - \underline{s}_0^T C_N^{-1}(\underline{s}_1-\underline{s}_0)}{\sqrt{(\underline{s}_1-\underline{s}_0)\, C_N^{-1}(\underline{s}_1-\underline{s}_0)}}\right]$$

sub. for λ'

$$= Q\left[\frac{\log\lambda + 1/2\,(\underline{s}_1-\underline{s}_0)\, C_N^{-1}(\underline{s}_1-\underline{s}_0)}{\sqrt{(\underline{s}_1-\underline{s}_0)\, C_N^{-1}(\underline{s}_1-\underline{s}_0)}}\right]$$

And:

Gaussian
zero mean
var same

$$P(H_1|H_1) = Pr\left[\underline{R}^T C_N^{-1}(\underline{s}_1-\underline{s}_0) > \lambda' \,|\, H_1\right]$$

$$= Q\left[\frac{\lambda' - \underline{s}_1^T C_N^{-1}(\underline{s}_1-\underline{s}_0)}{\sqrt{(\underline{s}_1-\underline{s}_0)\, C_N^{-1}(\underline{s}_1-\underline{s}_0)}}\right]$$

$$= Q\left[\frac{\log\lambda - 1/2\,(\underline{s}_1-\underline{s}_0)^T C_N^{-1}(\underline{s}_1-\underline{s}_0)}{\sqrt{(\underline{s}_1-\underline{s}_0)^T C_N^{-1}(\underline{s}_1-\underline{s}_0)}}\right]$$

etc...

Note: The features of the signals & noise which determine the performance are encapsulated by the quadratic form $(\underline{s}_1 - \underline{s}_0)^T C_N^{-1} (\underline{s}_1 - \underline{s}_0)$
generalized euclidean noise.

4.2 Partial Coherent Detection in White Gaussian Noise

Bandpass signals

maybe mistake in transcription

sent message $s_m(k) = a_m(kT_s)\cos(2\pi f_c T_s k + \phi_m(kT_s))$

- $a_m(kT_s)$ is the signal envelope (slowly varying compared to f_c)
- $\phi_m(kT)$ is the signal phase. (slowly varying same.)
- $\frac{1}{T_s}$ is the sampling rate such that $F_c T_s$ is sufficiently small. oversampling.
- F_c the carrier frequency

Hypotheses

$k = 0$ to $L-1$ composite hypothesis

received message $H_m: r(k) = a_m(kT_s)\cos(2\pi f_c T_s k + \phi_m(kT_s) + \Theta) + n(k)$

where:

$$f_{\theta|H}(\theta|H_m) = \frac{\exp\{\Omega_m \cos(\theta - \theta_0)\}}{2\pi I_0[\Omega_m]}, \quad -\pi < \theta < \pi$$
$$= 0 \quad \text{else}$$

θ_0
Tikkonov PDF
param Ω_m, θ_0
$\Omega_m \to 0$, uniform pdf
$\Omega_m \gg 1$, $\delta(\theta - \theta_0)$ pdf.

and: $I_0(x)$ is modified Bessel function first kind, order zero.

$$I_0(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{x\cos u}\,du$$

The noise is a sample of zero mean Gaussian vector $\underline{N}$

with $E[\underline{N}\,\underline{N}^T] = I$, $\underline{n} = [n(0), \ldots n(L-1)]^T$

$\boxed{\text{We have:}}$

$\underline{r} = [r(0) \ \ldots \ r(L-1)]^T$

$\underline{s}_m(\theta) = [s_m(0,\theta), \ \ldots \ s_m(L-1,\theta)]^T$

where;

$$s_m(k,\theta) = a_m(kT_s)\cos[2\pi f_c T_s k t + \phi_m(kT_s) + \theta]$$

$$H_m: f_{\underline{R}|H,B}(\underline{r}|H_m,\theta) = \frac{1}{(2\pi)^{L/2}} e\left[-1/2(\underline{r}-s_m(\theta))^T(\underline{r}-s_m(\theta))\right]$$

Under Bayesian framework:

with: P_0, P_1 are a prior probs, given

C_{ij} costs given, do not depend on θ

→ We can consider the Hypotheses:

in relation: see Q16 Chap. 3

$$\underline{H_m}: f_{\underline{R}|H,B}(\underline{r}|H_m) = \int_{-\infty}^{\infty} f_{\underline{R}|H,B}(\underline{r}|H_m,\theta)\, f_{B|H}(\theta|H_m)\, d\theta$$

→ problem: find solution of integral.

Calculating $f_{\underline{R}|H,B}(\underline{r}|H_m)$:

$$\underline{s}_m^T(\theta)\underline{s}_m(\theta) = \sum_{k=0}^{L-1} a_m^2(kT_s)\cos^2[2\pi f_c T_s k + \phi_m(kT_s) + \theta]$$

$$= 1/2\sum_{k=0}^{L-1} a_m^2(kT_s) + 1/2\underbrace{\sum_{k=0}^{L-1} a_m^2(kT_s)\cos[4\pi f_c T_s k + 2\phi_m(kT_s) + 2\theta]}_{\approx 0 \text{ for bandpass signal \& L large.}}$$

$$\to \boxed{\underline{s}_m^T(\theta)\underline{s}_m(\theta) = 1/2\sum_{k=0}^{L-1} a_m^2(kT_s) = E_m} \to \text{energy of } \underline{s}_m(\theta)$$

→ does not depend on θ

$$\underline{s}_m^T(\theta)\underline{r} = \underline{r}^T \underline{s}_m(\theta)$$

$$= \sum_{k=0}^{L-1} r(k)\, a_m(kT_s) \cos[2\pi f_c T_s k + \phi_m(kT_s) + \theta]$$

$$= A_c(\underline{r})\cos(\theta) - A_s(\underline{r})\sin(\theta)$$

where $A_c(\underline{r}) = \underline{r}^T \underline{s}_m(0)$

$A_s(\underline{r}) = \underline{r}^T \underline{s}_m(\pi/2)$

Substituting:

$$f_{\underline{R}|H,\Theta}(\underline{r}|H_m,\theta) = \frac{1}{(2\pi)^{L/2}} \exp[-1/2\,\underline{r}^T\underline{r}] \exp[-1/2\,E_m] \exp[A_c(\underline{r})\cos\theta - A_s(\underline{r})\sin\theta)$$

→ can now average over θ.

$$f_{\underline{R}|H}(\underline{r}|H_m) = \frac{1}{(2\pi)^{L/2}} \exp[-1/2\,\underline{r}^T\underline{r}] \exp[-1/2\,E_m]\, Z_m(\underline{r})$$

$$Z_m(\underline{r}) = \frac{1}{I_0(\Omega_m)} \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\{B_c(\underline{r})\cos\theta - B_s(\underline{r})\sin\theta\}\, d\theta$$

where: $B_c(\underline{r}) = A_c(\underline{r}) + \Omega_m \cos(\theta_0)$

$B_s(\underline{r}) = A_s(\underline{r}) - \Omega_m \sin(\theta_0)$

or $$Z_m(\underline{r}) = \frac{1}{I_0(\Omega_m)} \cdot \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left\{\sqrt{B_c^2(\underline{r}) + B_s^2(\underline{r})}\, \cos(\theta + \theta_R)\, d\theta\right.$$

where $\theta_B = \dfrac{B_s(\underline{r})}{B_c(\underline{r})}$

$$\therefore Z_m(\underline{r}) = \frac{I_0\left(\sqrt{B_c^2(\underline{r}) + B_s^2(\underline{r})}\right)}{I_0(\Omega_m)}$$

$$= \frac{I_0\left(\sqrt{[A_c(\underline{r}) + \Omega_m\cos\theta_0]^2 + [A_s(\underline{r}) - \Omega_m\sin(\theta_0)]^2}\right)}{I_0(\Omega_m)}$$

$$= \frac{I_0\left(\sqrt{A_c^2(\underline{r}) + A_s^2(\underline{r}) + 2\Omega_m[A_c(\underline{r})\cos\theta_0 - A_s(\underline{r})\sin\theta_0} + \Omega_m^2\right)}{I_0(\Omega_m)}$$

Look at 2 special cases:

For $\Omega_m = 0$: non-coherent case (no knowledge of phase)

$$z_m(\underline{r}) = I_0\left(\sqrt{A_c^2(\underline{r}) + A_s^2(\underline{r})}\right)$$

$\Rightarrow \theta_0$ not used.

For $\Omega_m \to \infty$: coherent case (known phase)

Use $I_0(x) \simeq \frac{e^x}{\sqrt{2\pi x}}, \quad x \gg 1$

$$z_m(\underline{r}) = \exp\left[A_c(\underline{r})\cos\theta_0 - A_s(\underline{r})\sin\theta_0\right]$$

Notice: $A_c(\underline{r})\cos\theta_0 - A_s(\underline{r})\sin\theta_0 = \underline{r}^T\left[\underline{s}_m(0)\cos\theta_0 - \underline{s}_m(\pi/2)\sin\theta_0\right]$

$= \underbrace{r^T \underline{s}_m(\theta_0)}_{\text{correlation of received signal with noiseless signal } \underline{s}_m(\theta_0)}$

$\to z_m(\underline{r}) = \exp\left[r^T \underline{s}_m(\theta_0)\right]$

$\to$ detection of known signal in noise.

In general $0 < \Omega_m < \infty$

$\to$ a combination of the coherent & non-coherent cases

$$z_m(\underline{r}) = \frac{1}{I_0(\Omega_m)} I_0\left(\sqrt{A_c^2(\underline{r}) + A_s^2(\underline{r}) + 2\Omega_m \underline{r}^T s_m(\theta_0) + \Omega_m^2}\right)$$

Notice weighing is by Ω_m.

distinguish between Gaussians of General mean, → correlation

[redacted]

4.3 The General Gaussian detection problem

$$H_m: f_{R|H_m}(\underline{r}|H_m) = \frac{1}{(2\pi)^{L/2}[\det(C_m)]^{1/2}} \exp\left[-1/2(\underline{r}-\underline{s}_m)^T C_m^{-1}(\underline{r}-\underline{s}_m)\right]$$

$$x \triangleq \Lambda_{1,0}(\underline{r}) = \frac{f_{R|H}(\underline{r}|H_1)}{f_{R|H}(\underline{r}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

$$\ln(\Lambda_{1,0}(\underline{r})) = \underline{r}^T Q \underline{r} + \underline{a}^T \underline{r} + b$$ non-homogeneous quadratic form

$$\left.\begin{aligned} Q &= 1/2\,(C_0^{-1} - C_1^{-1}) \\ \underline{a} &= [\underline{s}_1^T C_1^{-1} - \underline{s}_0 C_0^{-1}] \\ b &= 1/2 \ln \frac{\det(C_0)}{\det(C_1)} + \underline{s}_0^T C_0^{-1}\underline{s}_0 - \underline{s}_1^T C_1^{-1}\underline{s}_1 \end{aligned}\right\}$$ completely known

$$x \triangleq \ln \Lambda_{1,0}(\underline{R})$$

How to analyze error performance in a general way?

→ Use characteristic functions. (of $x \triangleq \Lambda_{1,0}(\underline{r})|H_m$

$$\boxed{\Phi_{x|H_m}(\omega) = E[e^{j\omega}|H_m]}$$ via Fourier Transform.

Note

$$\Phi_{x|H_m}(\omega) = \frac{\exp\left[-1/2(\underline{s}_m^T C_m^{-1}\underline{s}_m - j\omega 2b) + 1/2(\underline{s}_m + j\omega C_m \underline{a})^T (I - j2\omega Q C_m)^{-1} \cdot C_m^{-1}(\underline{s}_m + j\omega C_m \underline{a})\right]}{\sqrt{\det(I - j2\omega Q C_m)}}$$

$$P_F = P(H_1|H_0) = P(x > \log\lambda\,|H_0) = \int_{\log\lambda}^{\infty} f_{x|H_0}(x|H_0)\,dx$$

with $f(x|H_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_{x|H_m}(\omega)\, e^{-j\omega x}\, d\omega$

So: $P_F = \frac{1}{2\pi}\int_{\log\lambda}^{\infty}\int_{-\infty}^{\infty}\phi_{x|H_m}(\omega)\, e^{-j\omega x}\, d\omega\, dx$

<u>Use moment generating Function</u>

Let $M_{x|H_m}(z) = E\left[e^{zx}|H_m\right],\ z \in \mathbb{C}$

<u>Prop.</u>: $M_{x|H_m}(j\omega) = \phi_{x|H_m}(\omega)$

$M_{x|H_m}(z) = \phi_{x|H_m}(-jz)$

$$\text{So: } P_F = \frac{1}{2\pi j}\int_{\log\lambda}^{\infty}\int_{-j\infty}^{j\infty} M_{x|H_m}(z)\, e^{-zx}\, dz\, dx$$

$$= \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} M_{x|H_m}(z)\left[\int_{\ln\lambda}^{\infty} e^{-zx}\, dx\right] dz$$

$$= \frac{1}{2\pi j}\int_{-j\infty+\epsilon}^{j\infty+\epsilon} \frac{M_{x|H_m}(z)}{z}\, e^{-z\ln\lambda}\, dz$$

$\epsilon > 0$ is a small was introduced in order to move the path of integration away from the singularity $z=0$

<u>Another Method, due to "Imhof"</u>

$$P_F = 1/2 + \frac{1}{\pi}\int_0^{\infty} \frac{\mathrm{Im}\left[e^{-j\omega\ln\lambda}\,\phi_{x|H_0}(\omega)\right]}{\omega}\, d\omega$$

→ subtlety: where to truncate the numerical infinite integ

<u>Or, Find a bound which perhaps converges</u>

↘

4.4: Chernoff bounds to detection performance

Theorem 4.1 (Markov inequality)

let x be a non-negative R.V. $Pr\{x \geq 0\} = 1$

Then for any $a > 0$ we have: $Pr\{x \geq a\} < \frac{E(x)}{a}$

Proof:

$$Pr\{x \geq a\} = E[\phi_{[a,\infty]}(x)] \quad \text{where } \phi_{a,\infty} = \begin{cases} 1, & x \in [a,\infty] \\ 0, & \text{otherwise} \end{cases}$$

and $\frac{x}{a} = \begin{cases} \geq 1, & x \in [a,\infty] \\ \geq 0, & x \in [0,a] \end{cases}$

$$\therefore \quad \phi_{[a,\infty}(x) \leq \frac{x}{a}$$

$$\Rightarrow Pr\{x \geq a\} \leq E\left(\frac{x}{a}\right) = \frac{E(x)}{a}$$

$\square$

Consider the general test:

$$x(r) \underset{H_0}{\overset{H_1}{\gtrless}} \Theta$$

could be likelihood test, log-likelihood test, etc...

$x = 3$
$\Theta = 4$

$$P(H_1|H_0) = Pr(x > \Theta|H_0) = Pr[e^{sx} \geq e^{s\Theta} | H_0]$$ parameter $s > 0$ for all s

$e^{sx} > 0 \rightarrow$ apply Markov's inequality:

$$P(H_1|H_0) \leq e^{-s\Theta} E[e^{sx}|H_0]$$
$$= e^{-s\Theta} M_{x|H_0}(s) \qquad \text{for all } s > 0.$$

Define $\mu_{x|H_0}(s) = \ln M_{x|H_0}(s)$

the CGF, "Cumulant Generating Function"

Chernoff Bound

$$\boxed{P(H_1|H_0) \leq \exp\left[-s\theta + \mu_{x|H_0}(s)\right]} \quad \text{all } s>0$$

Similarly, Prob. of missing:

$$P(H_0|H_1) = Pr\{x < \theta | H_1\} = Pr[e^{sx} \geq e^{s\theta}|H_1] \quad s<0.$$

leads to:

Chernoff Bound

$$\boxed{P(H_0|H_1) \leq \exp\left[-s\theta + \mu_{x|H_1}(s)\right]} \quad \text{all } s<0$$

Theorem 4.2

Let x be a R.V. with Moment Generating function $M_x(s) = E[e^{sx}]$

and Cumulant Generating Function $\mu_x(s) = \ln M_x(s)$.

Then, $M_x(s)$ and $\mu_x(s)$ are convex functions.

→ take second derivative & check sign.

$$\frac{d}{ds}M_x(s) = \frac{d\,E[e^{sx}]}{ds} = E\left(\frac{de^{sx}}{ds}\right) = E\left(xe^{sx}\right)$$

$$\frac{d^2}{ds^2}M_x(s) = \frac{d\,E(xe^{sx})}{ds} = E\left(x^2e^{sx}\right] \geq 0 \quad \text{everywhere}$$

→ $M_x(s)$ is convex ∪

$$\frac{d}{ds}\mu_x(s) = \frac{1}{M_x(s)} \cdot \frac{d\,M_x(s)}{ds}$$

$$\frac{d^2}{ds}\mu_x(s) = \frac{M_x(s)\frac{d^2}{ds^2}M_x(s) - \left(\frac{dM_x(s)}{ds}\right)^2}{M_x^2(s)}$$

cauchy-schwarz inequality

$$\frac{dM_x(s)}{ds} = E^2[xe^{sx}] = E^2[xe^{sx/2}e^{sx/2}] \leq E[x^2e^{sx}] \cdot E[e^{sx}]$$

$$= \frac{d^2M_x(s)}{ds^2} \cdot M_x(s)$$

$$\rightarrow \frac{d^2\mu_x(s)}{ds^2} \geq 0 \quad \rightarrow \mu_x(s) \text{ is convex}$$

$$\rightarrow -\theta s + \mu_x(s) \text{ is convex.}$$

→ On any interval $[a,b]$, $-\theta s+\mu_x(s)$ has a minimum at $s_0=a$

$s_0=b$

or $b<s_0<a$

So the bounds on $P(H_0|H_1)$, $P(H_1|H_0)$ can be made maximum tight in the following way:

$$P(H_1|H_0)\leq \exp\left[\min_{s\geq 0}\left(-s\theta+\mu_{x|H_0}(s)\right)\right]$$

$$P(H_0|H_1)\leq \exp\left[\min_{s\leq 0}\left(-s\theta+\mu_{x|H_1}(s)\right)\right]$$

? →

$$\frac{d\mu_{x|H_m}(s)}{ds}=\dot{\mu}_{x|H_m}(s)$$

<u>Two possibilities</u>:

① $\dot{\mu}_{x|H_m}(s_m)=\theta$, such that $s_0>0$ or $s_1<0$

Then, the maximal tight bounds are:

$$P(H_1|H_0)\leq \exp\left[-s_0\dot{\mu}_{x|H_0}(s_0)+\mu_{x|H_0}(s_0)\right]$$

$$P(H_0|H_1)\leq \exp\left[-s_1\dot{\mu}_{x|H_1}(s_1)+\mu_{x|H_1}(s_1)\right]$$

② No such s_0, s_1 exist

Then, $-s\theta+\mu_{x|H_m}(s)$ achieves a minimum at $s=0$, and the bounds are useless

$(P(H_1|H_0)\leq \exp(0)=1 \ \ldots)$

6 $\boxed{\text{Estimation Theory}}$

Definition 6.1 An estimation rule (estimator) is a function from Z to $\mathcal{A}$ $\hat{a}(R): Z \to \mathcal{A}$

observation space   parameter space

Example 6.1

Suppose $r_i = a_i + w_i$, $i = 1$ to n

where w_i are realizations of i.i.d Gaussian RV $N(\underline{b}, 1)$ ← known vector

So:
$$f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a}) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi}} e^{-(r_i - b - a_i)^2/2}$$

Possible estimates of $\underline{a}$:

(i) $\hat{a}_i = r_i - b$, $i = 1, 2 \ldots M$ (linear)

(ii) $\hat{a}_i = (\sqrt{r_i} - \sqrt{b})^2$, $i = 1, 2 \ldots M$ (non-linear)

(iii) $\hat{a}_i = \dfrac{\sum_{i=1}^{M} r_i}{\left(\sum_{i=1}^{M} r_i\right)^2} (r_i - b)$ (not real-time)

→ will have different properties and performances.

Performance Measures

1) Bias: $\underline{b} = E[\underline{\hat{a}}(\underline{R}) - \underline{a}]$

→ First order performance measure.

2) Error covariance methods: $C[\underline{\hat{a}}(\underline{R})] = E\left[[\underline{\hat{a}}(\underline{R}) - E(\underline{\hat{a}}(\underline{R}))][\underline{\hat{a}}(\underline{R}) - E(\underline{\hat{a}}(\underline{R}))]^T\right]$

$$[C[\hat{a}(R)]]_{ij} = E\left[[\hat{a}_i(R) - E(\hat{a}_i(R))][\hat{a}_j(R) - E(\hat{a}_j(R))]^T\right]$$

→ a second order performance measure.

Two general models

① $\underline{a}$ is considered an unknown non-random parameter vector

② $\underline{a}$ is considered a realization of a random vector, and we know its statistics ("Bayesian Estimation")

6.2 estimation of unknown parameter

Performance Bounds

Let $\hat{\underline{a}}(\underline{r})$ be an unbiased estimator of $\underline{a}$, based on the observation $\underline{r}$, which is a realization of a random vector $\underline{r}$ with $f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})$, the likelihood function.

The score vectored function is defined as follows:

$$\underline{s}(\underline{r}|\underline{a}) = \frac{\partial \ln f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})}{\partial \underline{a}}$$

$$= \frac{1}{f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})} \cdot \frac{\partial f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})}{\partial \underline{a}}$$

class test upto here

Theorem 6.1

so that can exchange E() with ∫

Let $\underline{s}(\underline{r}|\underline{a})$ be the score of $f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})$ and $\underline{g}(\underline{r},\underline{a})$ be any well-behaved vectored function of $\underline{r}$ and $\underline{a}$, of appropriate dimension (?)

Then:

$$E[\underline{s}(\underline{R}|\underline{a})\underline{g}^T(\underline{R},\underline{a})] = \frac{\partial}{\partial \underline{a}} E[\underline{g}^T(\underline{R},\underline{a})] - E[\frac{\partial}{\partial \underline{a}} \underline{g}^T(\underline{R},\underline{a})]$$

Proof:

$$E[g^T(\underline{R},\underline{a})] = \int g^T(\underline{r},\underline{a}) f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})\, d\underline{r}$$

$$\frac{\partial}{\partial \underline{a}} E[g^T(\underline{R},\underline{a})] = \int \frac{\partial}{\partial \underline{a}} \left[g^T(\underline{r},\underline{a}) f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})\right] d\underline{r}$$

$$= \int \frac{\partial}{\partial \underline{a}} g^T(\underline{r},\underline{a}) f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})\, d\underline{r} + \int \frac{\partial f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})}{\partial \underline{a}} g^T(\underline{r},\underline{a})\, d\underline{r}$$

$$= E\left[\frac{\partial}{\partial \underline{a}} g^T(\underline{R},\underline{a})\right] + \int \underbrace{\frac{1}{f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})} \cdot \frac{\partial f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})}{\partial \underline{a}}}_{s(\underline{R}|\underline{a})} g^T(\underline{r},\underline{a}) f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})\, d\underline{r}$$

$$= E\left[\frac{\partial}{\partial \underline{a}} g^T(\underline{R},\underline{a})\right] + E[s(\underline{R}|\underline{a}) g^T(\underline{R},\underline{a})] \qquad \square$$

Consequences of theorem 6.1:

① $E[s(\underline{R}|\underline{a})] = 0$ (choosing $g = 1$)

② Let $\hat{\underline{a}}(\underline{R})$ be an unbiased estimator of $\underline{a}$.

Then: $$E[s(\underline{R}|\underline{a})\hat{\underline{a}}^T(\underline{R})] = \underbrace{\frac{\partial}{\partial \underline{a}} E[\hat{\underline{a}}(\underline{R})]}_{\underline{a} \text{ (since unbiased)}} - E\Big[\underbrace{\frac{\partial}{\partial \underline{a}} \hat{\underline{a}}^T(\underline{R})}_{\text{zero (since } \hat{\underline{a}} \text{ does not depend on } \underline{a})}\Big]$$

$$= \underline{a}$$

Definition

The covariance matrix of the score function is called the Fisher information matrix ("FIM") denoted by "J".

$$J = E[s(\underline{R}|\underline{a}) s^T(\underline{R}|\underline{a})]$$

$$= E\left[\frac{\partial \ln f_{\underline{R}|\underline{A}}(\underline{R}|\underline{a})}{\partial \underline{a}} \left(\frac{\partial \ln f_{\underline{R}|\underline{A}}(\underline{R}|\underline{a})}{\partial \underline{a}}\right)^T\right]$$

Using theorem 6.1 → with $g^T = s^T$

$$= -E\left[\frac{\partial}{\partial \underline{a}} s^T(\underline{R}|\underline{a})\right]$$

$$= -E\left[\frac{\partial}{\partial \underline{a}} \left(\frac{\partial \ln f_{\underline{R}|\underline{A}}(\underline{R}|\underline{a})}{\partial \underline{a}}\right)^T\right]$$

Definition (Notation)

Let A, B be symmetrical matrices of same dimensions.

Then $A \geq B$ means $[A-B] \geq 0$

$\underline{x}^T A \underline{x} \geq \underline{x}^T B \underline{x}$

And: $A > B$ means $[A-B] > 0$

$\underline{x}^T A \underline{x} > \underline{x}^T B \underline{x}$

Theorem 6.2 : Cramer-Rao

Let $\hat{\underline{a}}(\underline{r})$ be an unbiased estimator for the unknown parameter vector $\underline{a}$. Let $C(\hat{\underline{a}}(\underline{r}))$ be the covariance of this estimate.

Then $C(\hat{\underline{a}}(\underline{R})) \geq J^{-1}$, J^{-1} is assumed to exist.

Equality is satisfied if $\hat{\underline{a}}(\underline{r}) - \underline{a} = J^{-1}\underline{s}(\underline{r}|\underline{a})$

Proof

$$E[\underline{s}(\underline{R}|\underline{a})[\hat{\underline{a}}(\underline{R})-\underline{a}]^T] \overset{\text{Theom 6.1}}{=} \underbrace{E[\underline{s}(\underline{R}|\underline{a})\hat{\underline{a}}^T(\underline{R})]}_{=I} - E[\underline{s}(\underline{R}|\underline{a})]\underline{a}$$

$$E[J^{-1}\underline{s}(\underline{R}|\underline{a})[\hat{\underline{a}}(\underline{R})-\underline{a}]^T] = J^{-1}$$

$$\underline{x}^T E[J^{-1}\underline{s}(\underline{R}|\underline{a})[\hat{\underline{a}}(\underline{R})-\underline{a}]^T]\underline{x} = \underline{x}^T J^{-1}\underline{x}$$

$$E[\underline{x}^T J^{-1}\underline{s}(\underline{R}|\underline{a}) \cdot [\hat{\underline{a}}(\underline{R})-\underline{a}]^T\underline{x}] = \underline{x}^T J^{-1}\underline{x}$$

Cauchy-Schwartz $\rightarrow$ $$E[[\underline{x}^T J^{-1}\underline{s}(\underline{R}|\underline{a})]^2] \cdot E[[[\hat{\underline{a}}(\underline{R})-\underline{a}]^T\underline{x}]^2] \geq [\underline{x}^T J^{-1}\underline{x}]^2$$

$$E[\underline{x}^T J^{-1}\underline{s}(\underline{R}|\underline{a})\underline{s}^T(\underline{R}|\underline{a})J^{-1}\underline{x}] \cdot E[\underline{x}^T[\hat{\underline{a}}(\underline{R})-\underline{a}][\hat{\underline{a}}(\underline{R})-\underline{a}]^T\underline{x}] \geq [\underline{x}^T J^{-1}\underline{x}]^2$$

$$\underline{x}^T J^{-1}\underbrace{E[\underline{s}(\underline{R}|\underline{a})\underline{s}^T(\underline{R}|\underline{a})]}_{J}J^{-1}\underline{x} \cdot \underline{x}^T \underbrace{E[[\hat{\underline{a}}(\underline{R})-\underline{a}][\hat{\underline{a}}(\underline{R})-\underline{a}]^T]}_{C(\hat{\underline{a}}(\underline{R}))}\underline{x} \geq [\underline{x}^T J^{-1}\underline{x}]^2$$

$$\underline{x}^T J^{-1}\underline{x} \cdot \underline{x}^T C(\hat{\underline{a}}(\underline{R}))\underline{x} \geq [\underline{x}^T J^{-1}\underline{x}]^2$$

$$C(\hat{\underline{a}}(\underline{R})) \geq J^{-1}$$

equality: when condition holds.

$$C(\hat{\underline{a}}(\underline{R})) = E[(\hat{\underline{a}}(\underline{R})-\underline{a})(\hat{\underline{a}}(\underline{R})-\underline{a})^T]$$
$$= E[J^{-1}\underline{S}(\underline{R}|\underline{a})\,\underline{S}^T(\underline{R}|\underline{a})J^{-1}]$$
$$= J^{-1}E[\underline{S}(\underline{R}|\underline{a})\,\underline{S}^T(\underline{R}|\underline{a})]J^{-1} = J^{-1}$$

Definition

An unbiased estimator $\hat{\underline{a}}(\underline{R})$ such that $C[\hat{\underline{a}}(\underline{R})] = J^{-1}$ is called an efficient estimator. It may or may not exist.

Generalization of the Cramer-Rao Bound: Bhattacharyya class

Definition

Let $g(\underline{x})$ be a scalar function of a vector variable $\underline{x} = [x_1 \ldots x_n]$. Then, for any integer $k>0$, the k^{th} order gradient

$$\frac{\partial^k g(\underline{x})}{\partial \underline{x}^k} \triangleq \left[\frac{\partial^k g(\underline{x})}{\partial x_1^k}, \frac{\partial^k g(x)}{\partial x_2^k}, \ldots \frac{\partial^k g(x)}{\partial x_N^k} \right]$$

Provided that the derivatives exist.

Consider the estimation of the M dimensional vector $\underline{a}$, based on the n-dimensional observation $\underline{r}$, with likelihood function $f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})$.

Define the following

$$\underline{z}^0 = \hat{\underline{a}}(\underline{R}) - \underline{a}$$

$$\underline{z}^k = \frac{\partial^k \ln f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})}{\partial \underline{a}^k}, \quad k = 1, \ldots K, \text{ as long as derivatives exis}$$

Form the $M(K+1)$ dimensional vector:

$$\underline{z}^T = [\underline{z}^{0T}, \underline{z}^{1T}, \dots, \underline{z}^{KT}]$$

Consider the covariance matrix of $\underline{z}^T$:

or non-negative?

$$C_z = E[\underline{z}\underline{z}^T] \quad \rightarrow \text{symmetrical, positive definite}$$

It can be written in terms of $M \times M$ matrices

$$B^{ij} = E[\underline{z}^i \underline{z}^{jT}] \quad , \ i \text{ and } j = 0, 1, \dots, k.$$

<u>Note</u>: $B^{ij} = B^{ijT}$

$B^{ii} \neq B^{iiT}$

$$\rightarrow C_z = \begin{bmatrix} B^{00} & \cdots & B^{0K} \\ \vdots & B^{ij} & \vdots \\ B^{K0} & \cdots & B^{KK} \end{bmatrix}$$

Look at B^{ij}'s: $\quad B^{00} = E[\underline{z}^0 \underline{z}^{0T}] = \mathrm{Cov}(\hat{a}(\underline{R}))$

$$B^{10} = B^{01} = E[(\hat{\underline{a}} - \underline{a})\underline{s}^T(\underline{R}|\underline{a})] = I$$

$$k \geq 2 \ \underset{\text{component } ij}{[B^{k0}]_{ij} = [B^{0k}]_{ij}} = E\left[(\hat{a}_i(\underline{R}) - a_i)\frac{\partial^k \ln f_{\underline{R}|\underline{A}}(\underline{R}|\underline{A})}{\partial \underline{a}^k}\right]$$

$$= \frac{\partial^{k-1}}{\partial \underline{a}^{k-1}} E[\hat{a}_i(\underline{R}) s_j(\underline{R}|\underline{a})] - a_i \frac{\partial^{k-1}}{\partial \underline{a}^{k-1}} E[s_j(\underline{R}|\underline{a})]$$

From theorem 6.1:

$$E[\hat{a}_i(\underline{R}) s_j(\underline{R}|\underline{a})] = \begin{cases} 1, & i = j \\ 0 & \text{otherwise} \end{cases}$$

$$= [0]$$

$\therefore \ B^{k0} = B^{0k} = \vec{0}$ matrix

$$B^{11} = E[\underline{z}^1 \underline{z}^{1T}]$$

$$= E\left[\frac{\partial \ln f_{\underline{R}|\underline{A}}(\underline{R}|\underline{a})}{\partial \underline{a}} \left(\frac{\partial \ln f_{\underline{R}|\underline{A}}(\underline{R}|\underline{a})}{\partial \underline{a}}\right)^T\right]$$

$$= J$$

check

Therefore:

$$C_z = \begin{bmatrix} \mathrm{Cov}(\underline{a}) & U^T \\ U & B \end{bmatrix} \qquad U^T = [\, I \ \underbrace{[0]\ [0] \cdots [0]}_{k-1 \text{ matrices}} \,]$$

$$B = [B^{ij}] \quad i, j = 1, 2, \dots K$$

$$B^{11} = J$$

__Define the Schur complement of B in C_2:__

$$B^S = cov(\underline{a}) - U^T B^{-1} U$$

Denote $B^{-1} = \breve{B}$, and write it through submatrices of size $M \times M$

$\tilde{B}^{ij}$, $i,j = 1, \dots K$ check why?

Then, $B^S = cov(\underline{a}) - \tilde{B}^{11}$

Note: $C_2 \geq 0 \rightarrow B^S \geq 0 \rightarrow cov(\underline{a}) \geq \tilde{B}^{11}$

__Now, try to relate B^{11} to J.__

write: $B = \begin{bmatrix} J & B'^T \\ B' & B'' \end{bmatrix}$, $B'^T = [B^{1j}]$, $j = 2$ to k. sizes $M \times (k-1)M$

$B'' = [B^{ij}]$, $i,j = 2$ to k. $(K-1)M \times (k-1)M$.

$$\tilde{B}^{11} = J^{-1} + J^{-1} B'^T [B'' - B' J^{-1} B'^T]^{-1} B' J^{-1}$$

And the bound is

$$\boxed{cov(\underline{a}) \geq J^{-1} + J^{-1} B'^T (J^S)^{-1} B' J^{-1}}$$

$J^S = B'' - B' J^{-1} B'^T$ → Schur complement $B \geq 0 \rightarrow J^S \geq 0$

Show $J^{-1} B'^T (J^S)^{-1} B' J^{-1}$ is ~~pos.~~ non-negative definite, i.e. bound is tighter:

$$x^T J^{-1} B'^T (J^S)^{-1} B' J^{-1} \underline{x} = \underline{Y}^T \underbrace{(J^S)^{-1}}_{\text{non neg definite (known)}} \underline{Y} \geq 0$$

__Notes__

- $K=1$ generates the Cramer-Rao bound ($\tilde{B}^{11} = J^{-1}$)
- $K>1$, tighter bound
- If Cramer-Rao is shown tight, $J^{-1} B'^T (J^S)^{-1} B' J^{-1} = 0$.
  i.e. $J^{-1} B'^T (J^S)^{-1} B' J^{-1} = 0$ is necessary condition for efficient estimator.

$\boxed{\text{Example 6.2}}$

$$\underline{r} = a_1 + a_2 \underline{w} \quad \underline{w} \sim N(0,1)$$, estimate a_1 and a_2. what are bounds?

$$p_{R|A}(r|a) = \frac{1}{\sqrt{2\pi a_2^2}} \exp\left[-(r-a_1)^2 / 2a_2^2\right]$$

Find J

$$\ln f_{R|A}(r|a) = -1/2 \ln(2\pi) - \ln a_2 - \frac{1}{2a_2^2}(r-a_1)^2$$

$$\frac{\partial \ln f_{R|A}()}{\partial a_1} = \frac{1}{a_2^2}(r-a_1)$$

$$\frac{\partial \ln f_{R|A}()}{\partial a_2} = -\frac{1}{a_2} + \frac{1}{a_2^3}(r-a_1)^2$$

$$\frac{\partial^2 \ln f()}{\partial a_2 \partial a_1} = \frac{-2}{a_2^3}(r-a_1)$$

$$\frac{\partial^2 \ln f_{R|A}()}{\partial a_1^2} = -\frac{1}{a_2^2} \quad \bigg| \quad \frac{\partial^2 \ln f_{R|A}()}{\partial a_2^2} = +\frac{1}{a_2^2} - \frac{3}{a_2^4}(r-a_1)^2 \quad \bigg| \quad \frac{\partial^2 \ln f_{R|A}()}{\partial a_1 \partial a_2} = \frac{-2}{a_2^3}(r-a_1)$$

$$J = -E\left[\frac{\partial}{\partial \underline{a}}\left(\frac{\partial \ln f_{R|A}()}{\partial \underline{a}}\right)^T\right]$$

$$= E\begin{bmatrix} \frac{1}{a_2^2} & \frac{2}{a_2^3}(r-a_1) \\ \frac{2}{a_2^3}(r-a_1) & \frac{-1}{a_2^2} + \frac{3}{a_2^4}(r-a_1)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{a_2^2} & 0 \\ 0 & \frac{2}{a_2^2} \end{bmatrix}$$

$$J^{-1} = \begin{bmatrix} a_2^2 & 0 \\ 0 & a_2^2/2 \end{bmatrix}$$

Cramer Rao
$$\rightarrow \operatorname{Cov}([\hat{a}_1 \;\; \hat{a}_2]) \leq \begin{bmatrix} a_2^2 & 0 \\ 0 & a_2^2/2 \end{bmatrix}$$

For Ex: $[1, 0]\, C \begin{pmatrix} 1 \\ 0 \end{pmatrix} = E[(\hat{a}_1 - a_1)^2] \geq a_2^2$

$[0 \; 1]\, C \begin{bmatrix} 0 \\ 1 \end{bmatrix} = E[(\hat{a}_2 - a_2)^2] \geq 1/2\, a_2^2$

$[1 \; 0]\, C \begin{bmatrix} 0 \\ 1 \end{bmatrix} = E[(\hat{a}_1 - a_1)(\hat{a}_2 - a_2)] \geq 0$

Maximum Likelihood (ML) Estimation

When considered as a function of $\underline{a}$, then $f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})$ is called the likelihood function (i.e. fixed $\underline{r}$ = observed val

The maximum likelihood estimate is:

$$\boxed{\hat{\underline{a}}_{ML}(\underline{r}) = \underset{\underline{a} \in \mathcal{A}}{\arg\max}\ f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})}$$

Equivalently:

$$\boxed{\hat{\underline{a}}_{ML}(\underline{r}) = \underset{\underline{a} \in \mathcal{A}}{\arg\max}\ \underbrace{\ln f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})}_{\text{log likelihood function}}}$$

If $f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})$ is continuous in $\underline{a}$, then a necessary condition on $\hat{\underline{a}}_{ML}(\underline{r})$ is:

$$\boxed{\left[\frac{\partial}{\partial \underline{a}} \ln f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})\right]_{\hat{\underline{a}}_{ML}(\underline{r})} = 0}$$

(assuming interior point)
- → also check boundaries
- → sometimes no boundary

i.e. $\boxed{s(\underline{r}|\hat{\underline{a}}_{ML}) = \underline{0}}$

Note: similar expressions for discrete case.
- Note that when $\underline{r}$ is fixed, the likelihood function can be continuous in $\underline{a}$.

Theorem 6.3

If an efficient estimator exists, then it must be the maximum likelihood estimator.

Proof

For an ML estimator, we have

$$\underline{s}(\underline{r} \mid \hat{\underline{a}}_{ML}) = 0 .$$

Assume that we have an efficient estimator $\hat{a}$:

$$\hat{\underline{a}}(\underline{r}) - \underline{a} = J^{-1} \underline{s}(\underline{r} \mid \underline{a})$$

Then $\hat{\underline{a}}(\underline{r}) - \hat{\underline{a}}_{ML} = J^{-1} \underline{s}(\underline{r} \mid \hat{\underline{a}}_{ML}) = \underline{0}$

$$\rightarrow \hat{\underline{a}}(\underline{r}) = \hat{\underline{a}}_{ML}$$

□

Asymptotic Behavior of the ML estimator

Let $\underline{R}^i$, $i = 0, 1, \dots \ell-1$ be i.i.d observations, each with PDF $f(\underline{r} \mid \underline{a})$.

We want to form a maximum likelihood estimator of $\underline{a}$ based on realizations $\underline{r}^i$ of $\underline{R}^i$, $i = 0, \dots \ell-1$.

Form the likelihood function:

$$\boxed{f_{\underline{R}^0 \dots \underline{R}^{\ell-1} \mid \underline{A}}(\underline{r}^0, \dots \underline{r}^{\ell-1} \mid \underline{a}) = \prod_{i=0}^{L-1} f(\underline{r}^i \mid \underline{a})}$$

Log likelihood function:

$$\boxed{\ln f_{\underline{R}^0, \dots \underline{R}^{\ell-1} \mid \underline{A}}(\underline{r}^0, \dots \underline{r}^{\ell-1} \mid \underline{a}) = \sum_{i=0}^{L-1} \ln f(\underline{r}^i \mid \underline{a})}$$

The score function:

$$\boxed{\underline{s}(\underline{R}^0, \dots \underline{R}^{\ell-1} \mid \underline{a}) = \sum_{i=0}^{\ell-1} \underline{s}(\underline{R}^i \mid \underline{a})}$$

with $\boxed{\underline{s}(\underline{R}^i \mid \underline{a}) = \dfrac{\partial \ln f(\underline{R}^i \mid \underline{a})}{\partial \underline{a}}}$

Notice $\quad E[\underline{s}(\underline{R}^0, \dots \underline{R}^{\ell-1} | \underline{a})] = \underline{0}$

① <u>Using the central limit theorem:</u> $\quad$ a sum of covariances $\underline{J}$

$$\boxed{\underline{s}(\underline{R}^0, \dots, \underline{R}^{\ell-1} | \underline{a}) \xrightarrow[L \to \infty]{} N(\underline{0}, LJ)}$$

with $\quad J = E[\underline{s}(\underline{R}^i | \underline{a}) \underline{s}^T (\underline{R}^i | \underline{a})]$

② <u>Using the Taylor expansion:</u> (around $\hat{\underline{a}}_{ML}$)

$$\underline{0} = \underline{s}(\underline{r}^0, \dots \underline{r}^{\ell-1} | \hat{\underline{a}}_{ML}) \simeq$$

$$\underline{s}(\underline{r}^0, \dots \underline{r}^{\ell-1} | \underline{a}) + \frac{\partial}{\partial \underline{a}} \underline{s}^T(\underline{r}^0, \dots \underline{r}^{\ell-1} | \underline{a}) (\hat{\underline{a}}_{ML} - \underline{a})$$

$$\rightarrow \boxed{\underline{s}(\underline{r}^0, \dots \underline{r}^{\ell-1} | \underline{a}) = -\frac{\partial}{\partial \underline{a}} \underline{s}^T(\underline{r}^0, \dots \underline{r}^{\ell-1} | \underline{a}) (\hat{\underline{a}}_{ML} - \underline{a})}$$

$$\boxed{= -\sum_{i=0}^{L-1} \frac{\partial}{\partial \underline{a}} \underline{s}^T(\underline{r}^i | \underline{a}) (\hat{\underline{a}}_{ML} - \underline{a})}$$

<u>Law of Large numbers:</u>

$$\frac{1}{L} \sum_{i=0}^{L-1} \frac{\partial}{\partial \underline{a}} \underline{s}^T(\underline{r}^i | \underline{a}) \xrightarrow[L \to \infty]{\text{i.p.}} E\left[\frac{\partial}{\partial \underline{a}} \underline{s}^T(\underline{R}^i | \underline{a})\right] = -J$$

$$\underline{s}(\underline{r}^0, \dots \underline{r}^{\ell-1} | \underline{a}) \xrightarrow[L \to \infty]{} \underbrace{LJ(\hat{\underline{a}}_{ML} - \underline{a})}_{\text{from ①} \; \sim N(\underline{0}, LJ)}$$

<u>so</u>:

$$\boxed{(\hat{\underline{a}}_{ML} - \underline{a}) \sim N(\underline{0}, \tfrac{1}{L} J^{-1})}$$

or $\quad \boxed{\hat{\underline{a}}_{ML} \sim N(\underline{a}, \tfrac{1}{L} J^{-1})}$

To summarize,

<u>Proven result</u>: Properties of $\hat{a}_{ML}$ as $L \to \infty$

(i) $\hat{a}_{ML} \sim N(\underline{a}, \frac{1}{L} J^{-1})$

(ii) $\hat{a}_{ML}$ is asymptotically unbiased $(E(\hat{a}_{ML}) \to \underline{a})$

(iii) $\hat{a}_{ML}$ is asymptotically efficient $(\mathrm{cov}(\hat{a}_{ML}, \underline{a}) \to 0)$

(iv) $\hat{a}_{ML}$ is asymptotically consistent (i.e. tends to a deterministic value : $\hat{\underline{a}}_{ML} \to \underline{a}$ which is the value of the parameter).

$\boxed{\text{Ex} \,|\, 6.3}$: <u>Additive noise model.</u>

Suppose $\underline{r} = \underline{H}(\underline{a}) + \underline{w}$

where $\underline{r} = [r_0 \cdots r_N]^T$

$\underline{a} = [a_0 \cdots a_k]^T$

$\underline{H}(\underline{a}) = [H_1(\underline{a}), \cdots H_N(\underline{a})]^T$

$\underline{w}$ is a sample of an N-dimensional random vector with pdf $f_{\underline{w}}(\underline{w})$

Then:

① $f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a}) = f_{\underline{w}}(\underline{r} - \underline{H}(\underline{a}))$

$$\hat{\underline{a}}_{ML} = \arg\max_{a} f_{\underline{w}}(\underline{r} - \underline{H}(\underline{a}))$$

② <u>With Gaussian Noise</u> : $f_{\underline{w}}(\underline{w}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det R}} \exp(-1/2\, \underline{w}^T R^{-1} \underline{w})$ $\quad R = E(\underline{w}\underline{w}^T)$

$\to f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det R}} \exp(-1/2\, (\underline{r} - \underline{H}(\underline{a}))^T R^{-1} (\underline{r} - \underline{H}(\underline{a})))$

$\hat{a}_{ML} = \arg\max_{a}\, \underline{a}, (\underline{r} - \underline{H}(\underline{a})) R^{-1} (\underline{r} - \underline{H}(\underline{a}))$ (used log-likelihood funct.

③ $\overset{\boxed{\text{AWGN}}}{\text{white noise}}$: $R = \sigma^2 I$

$$\hat{\underline{a}}_{ML} = \arg\min_{a} \sum_{i=0}^{N-1} (r_i - H_i(\underline{a}))^2$$

$\boxed{\text{Ex. 6.4}}$

Let $R_0, \ldots R_N$ be independent Bernoulli RV's with unknown parameter a, i.e. $Pr[R_i=1]=a$, $Pr[R_i=0]=1-a$, $0<a<1$

Find ML estimator and Cramer-Rao bound. (& Compare)

$\boxed{\text{ML estimator}}$

① $P_{R_i|A}(r_i|a) = a^{r_i}(1-a)^{1-r_i}$, $i=0, \ldots N-1$

$$P_{\underline{R}|A}(\underline{r}|a) = \prod_{i=0}^{N-1} a^{r_i}(1-a)^{1-r_i}$$

$$\ln P_{\underline{R}|A}(\underline{r}|a) = \sum_{i=0}^{N-1}[r_i \ln(a) + (1-r_i)\ln(1-a)]$$

$$= \ln(a)\sum_{i=0}^{N-1} r_i + \ln(1-a)\sum_{i=0}^{N-1}(1-r_i)$$

$$S(\underline{r}|a) = \frac{\partial}{\partial a}\ln P_{\underline{R}|A}(\underline{r}|a)$$

$$= \frac{1}{a}\sum_{i=0}^{N-1} r_i - \frac{1}{1-a}\sum_{i=0}^{N-1}(1-r_i)$$

$$S(\underline{r}|\hat{a}_{ML}) = 0 \rightarrow (1-\hat{a}_{ML})\sum_{i=0}^{N-1} r_i = \hat{a}_{ML}\sum_{i=0}^{N-1}(1-r_i)$$

$$\rightarrow \boxed{\hat{a}_{ML} = \frac{1}{N}\sum_{i=0}^{N-1} r_i}$$

$\boxed{\text{Cramer-Rao Bound}}$

② $$\frac{\partial}{\partial a} S(\underline{r}|a) = -\frac{1}{a^2}\sum_{i=0}^{N-1} r_i - \frac{1}{(1-a)^2}\sum_{i=0}^{N-1}(1-r_i)$$

$$-E_r\left(\frac{\partial}{\partial a}S(\underline{r}|a)\right) = \frac{N}{a} + \frac{N}{1-a} = \frac{N}{a(1-a)}$$

$\rightarrow$ Cramer-Rao bound is: $\boxed{E[(\hat{a}-a)^2] \geq \frac{a(1-a)}{N}}$

$\boxed{\text{Compare ML estimate and Cramer-Rao Bound}}$

$$E[\hat{a}_{ML}] = a$$

$$E[(\hat{a}-a)^2] = \frac{1}{N^2} E\left[\left(\sum_{i=0}^{N-1}(R_i-a)\right)^2\right]$$ $\}$ R_i are independent.

$$= \frac{1}{N} E[(R_i-a)^2]$$

$$= \frac{1}{N}[a(1-a)^2 + (1-a)a^2]$$

$$= \frac{a(1-a)}{N} \quad \rightarrow \text{efficient estimator}$$

6.3 Estimation of random variables - Bayes framework

- Assume that $\underline{A}$ is a K-dimensional random vector with joint PDF $f_{\underline{A}}(\underline{a})$.
- Let $\underline{r}$ be the N-dimensional observation which is a realization of the random vector $\underline{R}$ with conditional PDF $f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})$.
- Let $\hat{\underline{a}}(\underline{r})$ be an estimator of $\underline{a}$ based on $\underline{r}$.
- The error vector is $\underline{\varepsilon} = \hat{\underline{a}}(\underline{r}) - \underline{a}$
- Define a cost function $c(\underline{\varepsilon})$ which is a scalar, non-negative function of $\underline{\varepsilon}$
- Define a conditional average cost when $\underline{A}=\underline{a}$

$$E_{\underline{R}}[c(\underline{\varepsilon})|\underline{A}=\underline{a}] = \int c(\underline{\varepsilon}) f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})\,d\underline{r}$$

Average cost:

$$\bar{C} = E_{\underline{A},\underline{R}}[c(\underline{\varepsilon})] = E_{\underline{A}}[E_{\underline{R}}[c(\underline{\varepsilon})|\underline{A}=\underline{a}]]$$

$$= \int\left\{\int c(\underline{\varepsilon}) f_{\underline{R}|\underline{A}}(\underline{r}|\underline{a})\,d\underline{r}\right\} f_{\underline{A}}(\underline{a})\,d\underline{a}$$

$$= \iint c(\underline{\varepsilon}) f_{\underline{R},\underline{A}}(\underline{r},\underline{a})\,d\underline{r}\,d\underline{a}$$

$$= \int\left\{\int c(\underline{\varepsilon}) \underbrace{f_{\underline{A}|\underline{R}}(\underline{a}|\underline{r})}_{\text{a posterior probability}}\,d\underline{a}\right\} f_{\underline{R}}\,d\underline{r}$$

a posterior probability : $f_{\underline{A}|\underline{R}}(\underline{a}|\underline{r}) = \dfrac{f_{\underline{A},\underline{R}}(\underline{a},\underline{r})}{f_{\underline{R}}(\underline{r})}$

$$= \int E[c(\underline{\varepsilon})|\underline{R}=\underline{r}]\, f_{\underline{R}}(\underline{r})\,d\underline{r}$$

a posterior cost given an observation $\underline{r}$

$$E[c(\underline{\varepsilon})|\underline{R}=\underline{r}] = \int c(\underline{\varepsilon}) f_{\underline{A}|\underline{R}}(\underline{a}|\underline{r})\,d$$

$$= \int c(\hat{\underline{a}}(\underline{r})-\underline{a})\, f_{\underline{A}|\underline{R}}(\underline{a}|\underline{r})\,d\underline{r}$$

$\boxed{\text{Theorem 6.4}}$

Let $c(\underline{\varepsilon})$ be the Bayes estimation cost function.

If: (i) $C(\underline{\varepsilon}) = C(-\underline{\varepsilon})$

(ii) $C(b\underline{\varepsilon} + (1-b)\underline{M}) \leq bC(\underline{\varepsilon}) + (1-b)C(\underline{M})$ (convex)

(iii) $f_{\underline{A}|\underline{R}}[\underline{a} + E(\underline{A}|\underline{r})] = f_{\underline{A}|\underline{R}}(E(\underline{A}|\underline{r}) - \underline{a})$ (centered at cond. mean)

Then, the estimator minimizing $\bar{C}$ is the conditional mean $\hat{A} = E[\underline{A}|\underline{r}]$.

Proof

$$\bar{C} = E[C(\underline{\hat{a}}(\underline{R}) - \underline{A})] = \int \overbrace{E[C(\underline{\hat{a}}(\underline{R}) - \underline{A})|\underline{r}]}^{\text{non-negative so}} f_{\underline{R}}(\underline{r})\,d\underline{r}$$

$$\min \bar{C} \Rightarrow \min E[C(\hat{a}(\underline{R}) - \underline{A}|\underline{r}] \quad (\text{“}\underline{r}\text{ fixed”})$$

let $\underline{z} \triangleq \underline{a} - E[\underline{A}|\underline{r}]$:

$$\Rightarrow E[C(\underline{\hat{a}}(\underline{R}) - \underline{A})|\underline{r}] = \int C(\underline{\hat{a}} - E[\underline{A}|\underline{r}] - \underline{z})\, f_{\underline{A}|\underline{R}}(\underline{z} + E[\underline{A}|\underline{r}])\,d\underline{z}$$

With $\underline{z}' = -\underline{z}$:

$$\Rightarrow E[C(\underline{\hat{a}}(\underline{R}) - \underline{A})|\underline{r}] = \int C(\underline{\hat{a}} - E[\underline{A}|\underline{r}] + \underline{z}')\, f_{\underline{A}|\underline{R}}(E(\underline{A}|\underline{r}) - \underline{z}')\,d\underline{z}'$$

Using (iii):

$$\Rightarrow E[C(\underline{\hat{a}}(\underline{R}) - A)|\underline{r}] = \int C(\underline{\hat{a}} - E[A|\underline{r}] + \underline{z}')\, f_{\underline{A}|\underline{R}}(\underline{z}' + E(A|\underline{r}))\,d\underline{z}' \quad (1)$$

Using (i):

$$\Rightarrow E[C(\underline{\hat{a}}(\underline{R}) - A)|\underline{r}] = \int C(-\underline{\hat{a}} + E[\underline{A}|\underline{r}] - \underline{z}')\, f_{\underline{A}|\underline{R}}(\underline{z}' + E[\underline{A}|\underline{r}])\,d\underline{z}'$$

$$= \int C(-\underline{\hat{a}} + E[\underline{A}|\underline{r}] + \underline{z})\, f_{\underline{A}|\underline{R}}(\underline{z} + E[\underline{A}|\underline{r}])\,d\underline{z}$$

$$\text{Use (iii)} \quad = \int C(-\underline{\hat{a}} + E[\underline{A}|\underline{r}] + \underline{z})\, f_{\underline{A}|\underline{R}}(\underline{z} + E[\underline{A}|\underline{r}])\,d\underline{z} \quad (2)$$

Averaging (1) & (2) (which are equal)

$$\Rightarrow E[C(\underline{\hat{a}}(\underline{R}) - \underline{A})|\underline{r}] = \frac{1}{2}\int [C(\underline{\hat{a}} - E[\underline{A}|\underline{r}] + \underline{z}) + C(-\underline{\hat{a}} + E(\underline{A}|\underline{r}) + \underline{z})]\, f_{\underline{A}|\underline{R}}(\underline{z} + E[\underline{A}|\underline{r}])\,d\underline{z}$$

$$\text{Use (ii)} \quad \geq \int C(\underline{z})\, f_{\underline{A}|\underline{R}}(\underline{z} + E[\underline{A}|\underline{r}])\,d\underline{z}$$

with equality if $\underline{\hat{a}} = E[\underline{A}|\underline{r}]$

□

Theorem 6.5

The estimator that minimizes the mean-square error is the conditional mean $E[\underline{A}|\underline{r}]$ (no conditions)

Proof

$$\bar{\mathcal{E}} = E[(\hat{\underline{a}}(\underline{R})-\underline{A})^T(\hat{\underline{a}}(\underline{R})-\underline{A})] = \int \{E[(\hat{\underline{a}}(\underline{r})-\underline{A})^T(\hat{\underline{a}}(\underline{R})-\underline{A})|\underline{r}]\, f_{\underline{R}}(\underline{r})dz$$

minimize $\bar{\mathcal{E}}$ $\equiv$ minimize $E[(\hat{\underline{a}}(\underline{r})-\underline{A})^T(\hat{\underline{a}}(\underline{R})-\underline{A})|\underline{r}]$

$$E[(\hat{\underline{a}}(\underline{r})-\underline{A})^T(\hat{\underline{a}}(\underline{r})-\underline{A})|\underline{r}] = \int (\hat{\underline{a}}(\underline{r})-\underline{a})^T(\hat{\underline{a}}(\underline{r})-\underline{a})\, f_{\underline{A}|\underline{R}}(\underline{a}|\underline{r})d\underline{a}$$

$$\rightarrow \frac{\partial}{\partial \underline{a}} E[(\hat{\underline{a}}(\underline{r})-\underline{A})^T(\hat{\underline{a}}(\underline{r})-\underline{A})] = 2\int (\hat{\underline{a}}(\underline{r})-\underline{a})\, f_{\underline{A}|\underline{R}}(\underline{a}|\underline{r})d\underline{a} \;\underset{\text{to minimize.}}{=0}$$

$$\rightarrow \hat{\underline{a}} \underbrace{\int f_{\underline{A}|\underline{R}}(\underline{a}|\underline{r})d\underline{a}}_{1} = \underbrace{\int \underline{a}\, f_{\underline{A}|\underline{R}}(\underline{a}|\underline{r})d\underline{a}}_{E(\underline{A}|\underline{r})}$$

$$\rightarrow \boxed{\hat{\underline{a}} = E(\underline{A}|\underline{r})}$$

$$\begin{bmatrix} a & b & c \end{bmatrix}\begin{bmatrix} d & e & f \\ g & h & i \\ j & k & l \end{bmatrix} = \begin{bmatrix} ad & ae & af \\ bg & bh & bi \\ cj & ck & cl \end{bmatrix}$$

$$tr[M(RA^2R^T + \sigma^2 R)M^T - ARM^T - MRA]$$

$$= tr[M(RA^2R)] + \sigma^2 tr(MRM^T) - 2tr[ARM^T] \cancel{- tr[MRA]} \quad A^T R M^T$$

$$= tr[(A^2R)^T(MR)^T] + \sigma^2 tr(MRM^T) - 2tr[ARM^T]$$

$$= tr[AR\,R^T M^T + \sigma^2 MRM^T - 2ARM^T]$$

(A)

$$= tr[(AR^2 + \sigma^2 MR - 2AR)M^T] = tr[\overbrace{(AR^2 - 2AR)}^{C}M^T + \sigma^2 M\overbrace{R}^{B}M^T]$$

$$\begin{bmatrix} \; \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = [\;\; = tr[(AR + \sigma^2 M - 2A)RM^T] \quad = tr[CM^T + MBM^T]$$

$$\cancel{\begin{bmatrix} e & f \\ g & h \end{bmatrix}} \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} [a+d] \quad = tr[\{\underbrace{(R-2I)A}_{\triangleq C} + \sigma^2 M\}RM^T] \quad \begin{bmatrix} c_{11} & c_{1n} \\ c_{n1} & c_{mn} \end{bmatrix}\overset{M^T}{\begin{bmatrix} m_{11} & m_{n1} \\ m_{1n} & m_{nn} \end{bmatrix}}$$

$$[e\; f]\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = [\;] \quad tr[(C + \sigma^2 M)RM^T] \quad \text{trace of first:}$$

$$[\;] \quad a \quad = tr[(\quad \int_0^T x^2(t)dt = \Sigma \quad (c_{11}m_{11} + \dots + c_{1n}m_{1n}) + (c_{21}m_{21} + \dots +$$

E

$C_x(t,u)$ which satisfy

MBM^T

$(MB)^T M$

$B^T M^T M$

7.2 Mercer: With the same notation as in theorem 7.1, we have ~~the~~

$$C_x(t,u) = \sum_{i=1}^{\infty} \lambda_i \phi_i(t)\phi_i(u)$$

Similar to eigenvalue decomposition of matrices $A^H A$