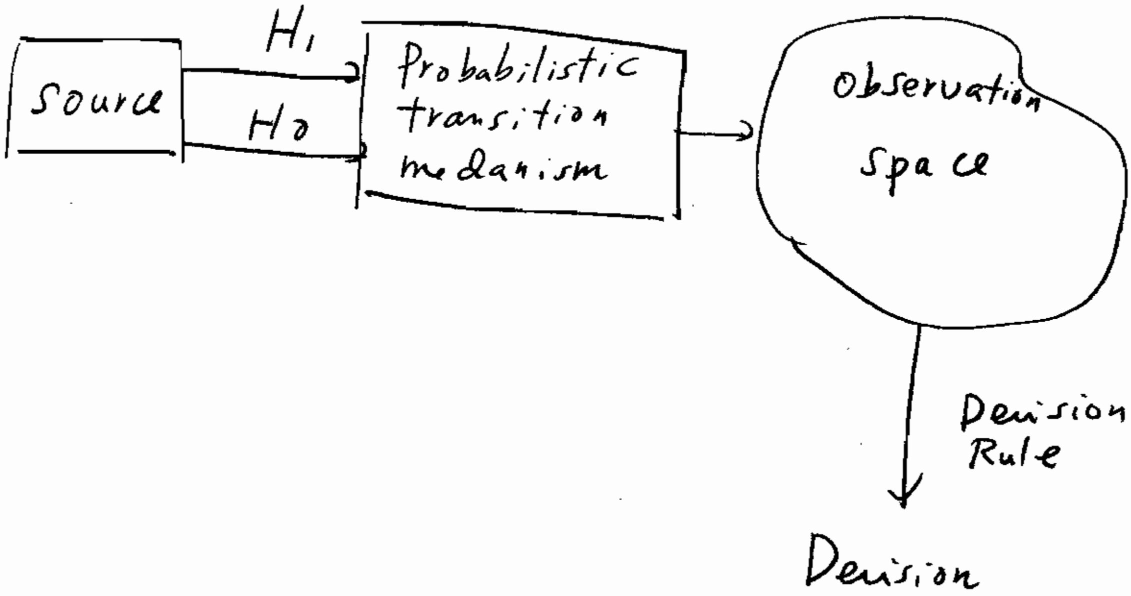
Chapter 2. Classical Detection and Estimation Theory



The basic components of a simple decision theory (binary hypothesis) are shown above.

The source (the nature, machine, etc.) generates an output. We refer the outcomes as the hypotheses and label them H_0 and H_1.

A usual case for this binary hypothesis test is to detect signal or noise, such that

H_0: n (noise)

H_1: $s+n$ (signal plus noise)

2.2 Simple Binary Hypothesis Tests

We assume that the observation space corresponds to a set of N observations: $r_1, r_2, r_3, \ldots, r_N$. Thus each set can be thought of as a point in an N-dimensional space and can be denoted by a vector $\vec{r}$:

$$\vec{r} \triangleq \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}$$

The probabilistic transition mechasim generates points in accordance with the two known conditional probability densities $f_{\vec{r}|H_1}(\vec{r}|H_1)$ and $f_{\vec{r}|H_0}(\vec{r}|H_0)$. The object is to use this information to develop a suitable decision rule.

2.2.1 Decision Criteria

In the binary hypothesis problem we know that either H_0 or H_1 is true.

Thus, each time the experimat is conducted, and we will have one of the four following outcomes:

1. H_0 true, choose H_0
2. H_0 true, choose H_1 — correct
3. H_1 true, choose H_1
4. H_1 true, choose H_0 — incorrect

X. Bayes Criterion

To evaluate how good the decision rule to be designed is, we can apply the Bayes criterion. A Bayes test is based on the two assumptions. The first is that the source outputs (a series of H_0 and H_1) are governed by the probility assigments which are denoted as P_0 and P_1, respectively.

In other wards, we say that the a priori probabilities are given. These a priori probabilities represent the required information about the source before any experimat is conducted. The second assumption is that a cost is assigned to each of the four aforementioned outcomes and we usually denote the costs as C_{00}, C_{10}, C_{11}, C_{01}, accordingly

$$C_{xy}$$

the chosen hypothesis

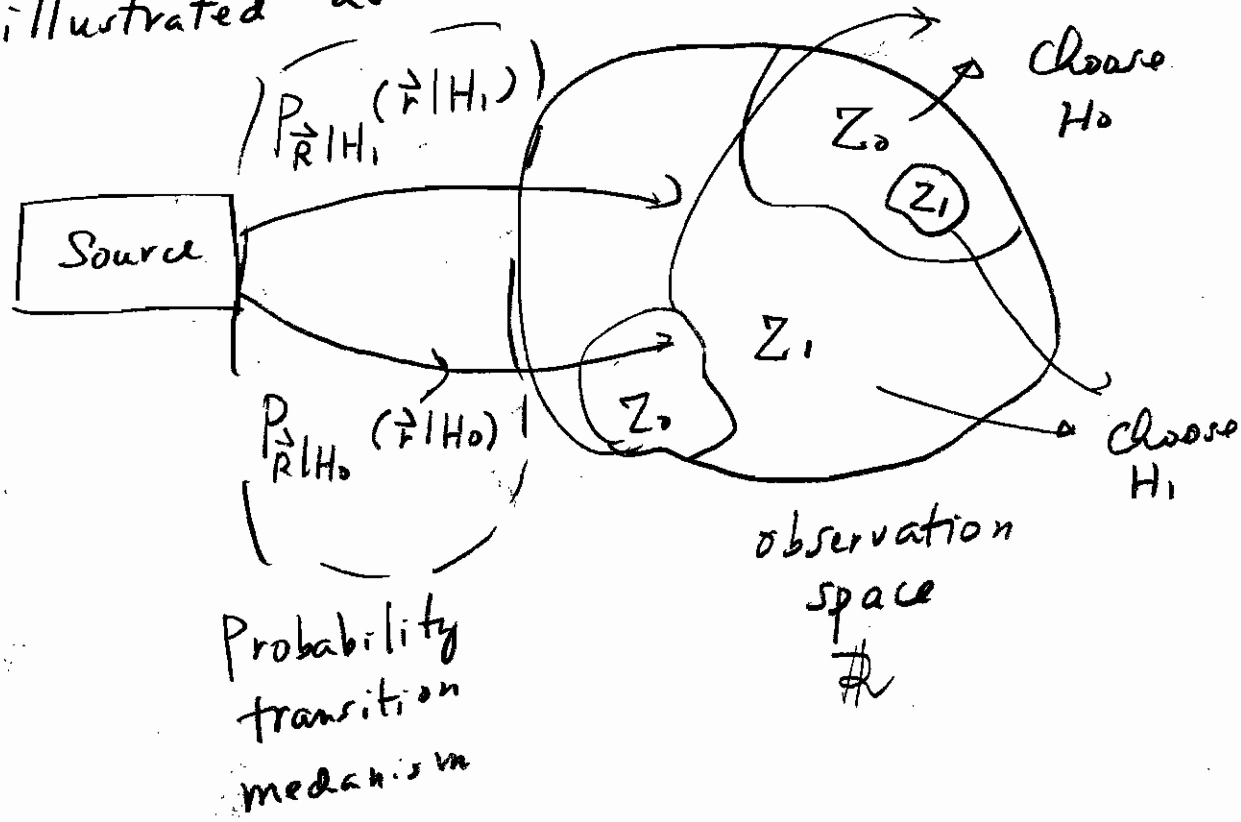
the true hypothesis

Thus, we can evaluate any decision rule based on the average cost. The average (or expected) cost can be written as the risk R such that

$$R = C_{00}\, P_0 \quad P\{\text{chosen } H_0 \mid H_0 \text{ true}\}$$
$$+ C_{10}\, P_0 \quad P\{\text{chosen } H_1 \mid H_0 \text{ true}\}$$

$$+ C_{11} P_1 \, P\{\text{chosen } H_1 \mid H_1 \text{ true}\}$$
$$+ C_{01} P_1 \, P\{\text{chosen } H_0 \mid H_1 \text{ true}\}$$

Assume that the observation $\vec{r}$ spans a space $\mathcal{R}$. For example, an N-dimensional vector space. Then, the decision rule is related to the allocation of the different regions pertaining to the individual hypotheses H_0, H_1, or the establishment of the decision boundaries between H_0 and H_1. It can be illustrated as below:



We can see from the illustration that Z_0, Z_1 specify the decision regions for H_0 and H_1, respectively. Thus, we can carefully design suc Z_0, Z_1 to minimize the risk R. Given C_{ij}, $i,j \in \{0,1\}$ and P_i, $i=0,1$, we can write the risk R as

$$R \triangleq C_{00} P_0 \int_{Z_0} P_{\vec{R}|H_0}(\vec{r}\,|H_0)\, d\vec{r}$$

$$+ C_{10} P_0 \int_{Z_1} P_{\vec{R}|H_0}(\vec{r}\,|H_0)\, d\vec{r}$$

$$+ C_{11} P_1 \int_{Z_1} P_{\vec{R}|H_1}(\vec{r}\,|H_1)\, d\vec{r}$$

$$+ C_{01} P_1 \int_{Z_0} P_{\vec{R}|H_1}(\vec{r}\,|H_1)\, d\vec{r}.$$

Next, we can assume that the cost of a incorrect decision is higher than that of a correct decision, i.e.,

$$C_{10} > C_{00}$$

$$C_{01} > C_{11}.$$

Now, we need to apply the Bayes test to choose the optimal decision regions Z_0, Z_1

in such a manner that the risk R is minimized.

Assume that the rejection of decision is not allowed and we have to assign each $\vec{r} \in Z$ to Z_0 or Z_1 exclusively. Thus, $Z = Z_0 \cup Z_1 \triangleq Z_0 + Z_1$,

$Z - Z_0 \triangleq Z \setminus Z_0$, $Z - Z_1 \triangleq Z \setminus Z_1$.

Then,

$$R = P_0 C_{00} \int_{Z_0} P_{\vec{R}|H_0}(\vec{r}|H_0)\, d\vec{r} +$$

$$P_0 C_{10} \int_{\underbrace{Z-Z_0}_{Z_1}} P_{\vec{R}|H_0}(\vec{r}|H_0)\, d\vec{r} +$$

$$P_1 C_{01} \int_{Z_0} P_{\vec{R}|H_1}(\vec{r}|H_1)\, d\vec{r} +$$

$$P_1 C_{11} \int_{\underbrace{Z-Z_0}_{Z_1}} P_{\vec{R}|H_1}(\vec{r}|H_1)\, d\vec{r}$$

Since $\int_Z P_{\vec{R}|H_0}(\vec{r}|H_0)\, d\vec{r} = \int_Z P_{\vec{R}|H_1}(\vec{r}|H_1)\, d\vec{r} = 1$,

we get

$$R = P_0 C_{10} + P_1 C_{11} + \int_{Z_0} \left\{ \left[P_1 (C_{01} - C_{11}) P_{\vec{R}|H_1}(\vec{r}|H_1) \right] - \left[P_0 (C_{10} - C_{00}) P_{\vec{R}|H_0}(\vec{r}|H_0) \right] \right\} d$$

Since $P_1 (C_{01} - C_{11}) \, p_{\vec{R}|H_1} (\vec{r}|H_1) \geq 0$

$P_0 (C_{10} - C_{00}) \, p_{\vec{R}|H_0} (\vec{r}|H_0) \geq 0$,

for each observation vector $\vec{r}$, we shall assign $\vec{r}$ to Z_1 if

$$P_1 (C_{01} - C_{11}) \, p_{\vec{R}|H_1} (\vec{r}|H_1)$$
$$\geq P_0 (C_{10} - C_{00}) \, p_{\vec{R}|H_0} (\vec{r}|H_0).$$

Or, in other words, the optimal decision rule is

$$\frac{p_{\vec{R}|H_1} (\vec{r}|H_1)}{p_{\vec{R}|H_0} (\vec{r}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0 (C_{10} - C_{00})}{P_1 (C_{01} - C_{11})}.$$

The decision boundaries are specified by the above equality. We can denote the decision boundary function as the likelihood function $\Lambda(\vec{r})$ such that

$$\Lambda(\vec{r}) \triangleq \frac{p_{\vec{R}|H_1} (\vec{r}|H_1)}{p_{\vec{R}|H_0} (\vec{r}|H_0)}$$

The constant consisting of costs and a priori probabilities can be denoted as the threshold of the hypothesis test such that

$$\eta \triangleq \frac{P_0 (C_{10} - C_{00})}{P_1 (C_{01} - C_{11})}$$

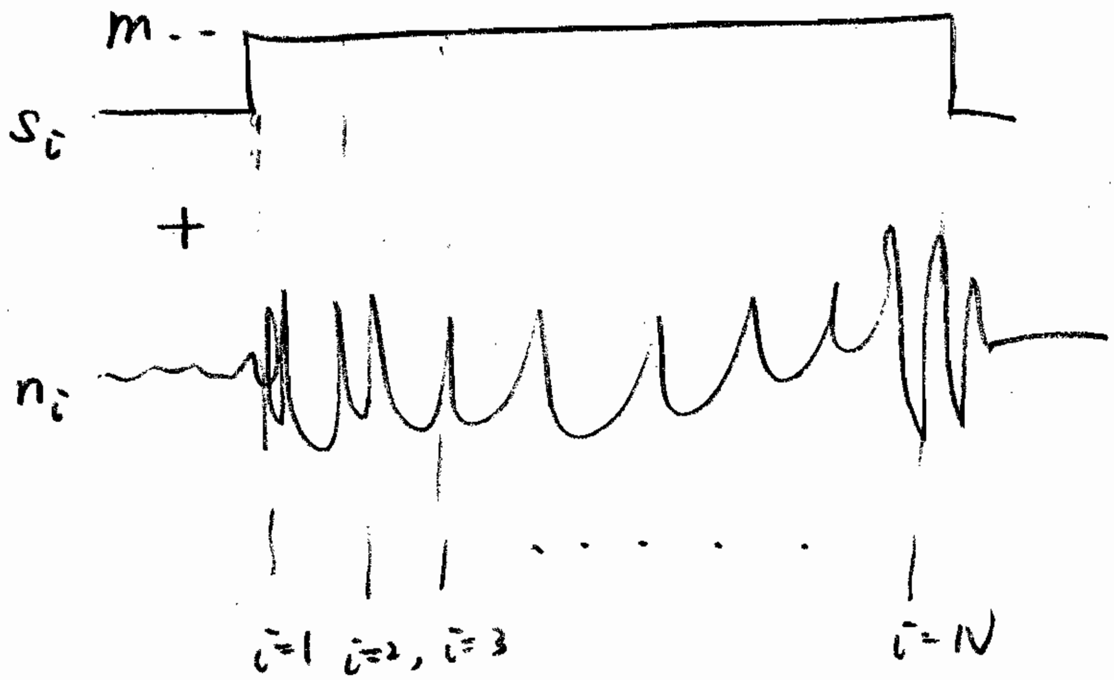
Thus, the Bayes criterion (test) for such a binary hypothesis leads to the likelihood ratio test (LRT) such that

$$\Lambda(\vec{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta.$$

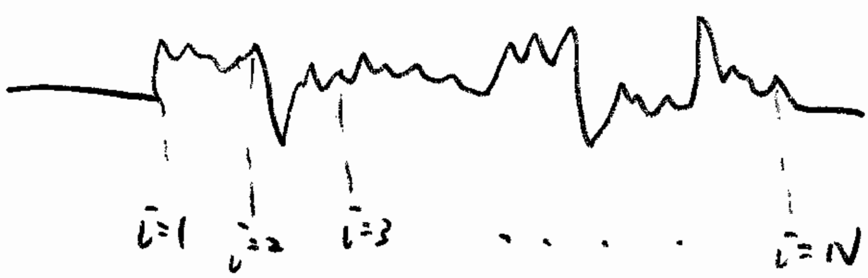
For some particular PDF $f_{\vec{R}|H_0}(\vec{r}|H_0)$, $f_{\vec{R}|H_1}(\vec{r}|H_1)$ such as Gaussian, we can simplify the LRT as the log-likelihood ratio test such that

$$\log(\Lambda(\vec{r})) \underset{H_0}{\overset{H_1}{\gtrless}} \log(\eta).$$

Example: The experiment is conducted to take several values from the constant binary pulse corrupted by the additive Gaussian noise with zero mean and variance σ^2. We acquire the sample values r_i, $i = 1, 2, \ldots, N$ to determine whether a binary "1" or "0" is sent.



$r_i = S_i + n_i$



We can construct a binary hypothesis test as

$$H_1 : r_i = m + n_i , \quad i = 1, 2, \ldots, N$$
$$H_0 : r_i = n_i , \quad i = 1, 2, \ldots, N$$

and $f_{N_i} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{n_i^2}{2\sigma^2}\right)$

i.i.d.

The PDF of r_i can be formulated as

$$f_{R_i|H_1}(r_i | H_1) = f_{N_i}(r_i - m)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r_i - m)^2}{2\sigma^2}\right)$$

$$f_{R_i|H_0}(r_i | H_0) = f_{N_i}(r_i)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{r_i^2}{2\sigma^2}\right)$$

Because n_i are statistically independent of each other, the joint PDF of the entire set of observations $\vec{r}$ is

$$f_{\vec{R}|H_1}(\vec{r}|H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(r_i - m)^2}{2\sigma^2}\right]$$

$$f_{\vec{R}|H_0}(\vec{r}|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{r_i^2}{2\sigma^2}\right]$$

We have the likelihood ratio as

$$\Lambda(\vec{r}) = \frac{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(r_i-m)^2}{2\sigma^2}\right]}{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{r_i^2}{2\sigma^2}\right]}$$

The log-likelihood ratio is

$$\log \Lambda(\vec{r}) = \frac{m}{\sigma^2} \sum_{i=1}^{N} r_i - \frac{Nm^2}{2\sigma^2}$$

Thus, the likelihood ratio test is

$$\frac{m}{\sigma^2} \sum_{i=1}^{N} r_i - \frac{Nm^2}{2\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \log \eta$$

Or equivalently,

$$\sum_{i=1}^{N} r_i \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma^2}{m} \ln \eta + \frac{Nm}{2} \triangleq \gamma$$

Example: The observations consist of a set of N values: $r_1, r_2, r_3, \ldots, r_N$.

Under H_1, each r_i has a variance σ_1^2.
Under H_0, each r_i has a variance σ_0^2.

Then, we can construct the two PDFs:

$$f_{\vec{R}|H_0}(\vec{r}\,|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left(-\frac{r_i^2}{2\sigma_0^2}\right)$$

$$f_{\vec{R}|H_1}(\vec{r}\,|H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left(-\frac{r_i^2}{2\sigma_1^2}\right)$$

Thus, we can form the log-likelihood as

$$\log \Lambda(\vec{r}) = \log\left[\frac{f_{\vec{R}|H_1}(\vec{r}\,|H_1)}{f_{\vec{R}|H_0}(\vec{r}\,|H_0)}\right]$$

$$= \frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum_{i=1}^{N} r_i^2$$

$$+ N\log\left(\frac{\sigma_0}{\sigma_1}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \log \eta$$

We can define the sufficient statistics as the sum of the squares of the observations

$$l(\vec{r}) = \sum_{i=1}^{N} r_i^2,$$

If $\sigma_1^2 > \sigma_0^2$, we have a mathematically equivalent test as

$$l(\vec{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\left(\log\eta - N\log\frac{\sigma_0}{\sigma_1}\right)$$

$$\triangleq \gamma$$

If $\sigma_1^2 < \sigma_0^2$, we have the reverse test as

$$l(\vec{r}) \underset{H_1}{\overset{H_0}{\gtrless}} \frac{2\sigma_0^2\sigma_1^2}{\sigma_0^2 - \sigma_1^2}\left(N\log\frac{\sigma_0}{\sigma_1} - \log\eta\right)$$

$$\triangleq \gamma'$$

If we set the costs $C_{00} = C_{11} = 0$ and $C_{10} = C_{01} = 1$, the risk is simplified as

$$R = P_0 \int_{Z_1} f_{\vec{R}|H_0}(\vec{r}|H_0)\, d\vec{r}$$

$$+ P_1 \int_{Z_0} f_{\vec{r}|H_1}(\vec{r}|H_1)\, d\vec{r}$$

Then, the Bayes test is given by

$$\log\Lambda(\vec{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \log\left(\frac{P_0}{P_1}\right) = \log(P_0) - \log(1-P_0).$$

We can denote the following parameters:

$$P_F \triangleq \int_{Z_1} f_{\vec{R}|H_0}(\vec{r}\,|H_0)\,d\vec{r}$$ (Probability of a false alarm)

$$P_D \triangleq \int_{Z_1} f_{\vec{R}|H_1}(\vec{r}\,|H_1)\,d\vec{r}$$ (Probability of detection)

$$P_M \triangleq \int_{Z_0} f_{\vec{R}|H_1}(\vec{r}\,|H_1)\,d\vec{r}$$ (Probability of miss detection)

$$= 1 - P_D$$

$$R = P_0 C_{10} + P_1 C_{11} + P_1 (C_{01} - C_{11}) P_M - P_0 (C_{10} - C_{00})(1 - P_F)$$

Since $P_0 = 1 - P_1$,

$$R_B(P_1) = C_{00}(1 - P_F) + C_{10} P_F + P_1 \left[(C_{11} - C_{00}) + (C_{01} - C_{11}) P_M - (C_{10} - C_{00}) P_F \right].$$

If all of the costs and a priori probabilities are known, we can determine a Bayes test $R_B(P_1)$, which is a function of P_1.

Since $\eta = \dfrac{(1-P_1)(C_{10} - C_{00})}{P_1 (C_{01} - C_{11})}$

$P_1 \uparrow \Rightarrow \eta \downarrow \Rightarrow P_F \uparrow, \ P_M \downarrow$

※ Neyman-Pearson Tests:

In many physical situations it is difficult to assign realistic costs or a priori probabilities. A simple procedure to by-pass this difficulty is to work with the conditional probabilities P_F and P_D. In general, we should make P_F as small as possible and P_D as large as possible. For most problems of practical importance there are conflicting objectives. An obvious criterion is to constrain one of the probabilities and maximize or minimize the other. A specific statement of this criterion (Neyman-Pearson Criterion) is as follows:

Neyman-Pearson Criterion.

Constrain $P_F = \alpha' \leq \alpha$ and design a test (not Bayes test) to maximize P_D (or minimize P_M) under this given constraint.

The solution is obtained easily by using Lagrange multipliers. We can construct a Lagrange multiplier function as

$$F = P_M + \lambda \left[P_F - \alpha' \right],$$

or

$$F = \underbrace{\int_{Z_0} f_{\vec{R}|H_1}(\vec{r}|H_1)\, d\vec{r}}_{\text{objective function}} + \lambda \underbrace{\left[\int_{Z_1} f_{\vec{R}|H_0}(\vec{r}|H_0)\, d\vec{r} - \alpha' \right]}_{\text{constraint function}}$$

Clearly, if $P_F = \alpha'$, then $\min\{F\} = \min\{P_M\}$

Since $$\int_{Z_0} f_{\vec{R}|H_0}(\vec{r}|H_0)\,d\vec{r} + \int_{Z_1} f_{\vec{R}|H_1}(\vec{r}|H_1)\,d\vec{r} = 1,$$

we get

$$F = \lambda(1-\alpha') + \int_{Z_0}\left[f_{\vec{R}|H_1}(\vec{r}|H_1) - \lambda f_{\vec{R}|H_0}(\vec{r}|H_0)\right]d\vec{r}$$

Obviously, any $\lambda > 0$, an LRT will minimize F, which is

$$\frac{f_{\vec{R}|H_1}(\vec{r}|H_1)}{f_{\vec{R}|H_0}(\vec{r}|H_0)} < \lambda,$$ assign $\vec{r}$ to Z_0 or choose h

Or
$$\Lambda(\vec{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

To satisfy the constraint function, we have to choose λ so that $P_F = \alpha'$.

Denote a new variable

$$\Lambda = \frac{f_{\vec{R}|H_1}(\vec{r}\,|H_1)}{f_{\vec{R}|H_0}(\vec{r}\,|H_0)}$$

$\downarrow$

a function of $\vec{r}$
scalar

Thus,

$$P_F = \int_{\lambda}^{\infty} f_{\Lambda|H_0}(\Lambda\,|\,H_0)\, d\Lambda = \alpha'$$

$\Rightarrow$ determine the required threshold λ.

Observe that $\lambda \downarrow$, $P_F \uparrow$, increasing the region Z_1 and $P_D \uparrow$

It is noted that $\Lambda \geq 0$, hence $\lambda \geq 0$.

The smallest possible λ ($\lambda = 0$) will be corresponding to the largest P_F or

the $P_F = \alpha' \leq \alpha$ is bounded.

※ Summary of Bayes and Neyman-Pearson Tests:

① Given the observation vector $\vec{r} \in Z$, we can develop the likelihood ratio

$$\Lambda(\vec{r}) = \frac{f_{\vec{R}|H_1}(\vec{r}|H_1)}{f_{\vec{R}|H_0}(\vec{r}|H_0)}$$

② Then build an LRT such that

$$\Lambda(\vec{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

③ Determine the threshold λ using

(i) Bayes optimal threshod

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

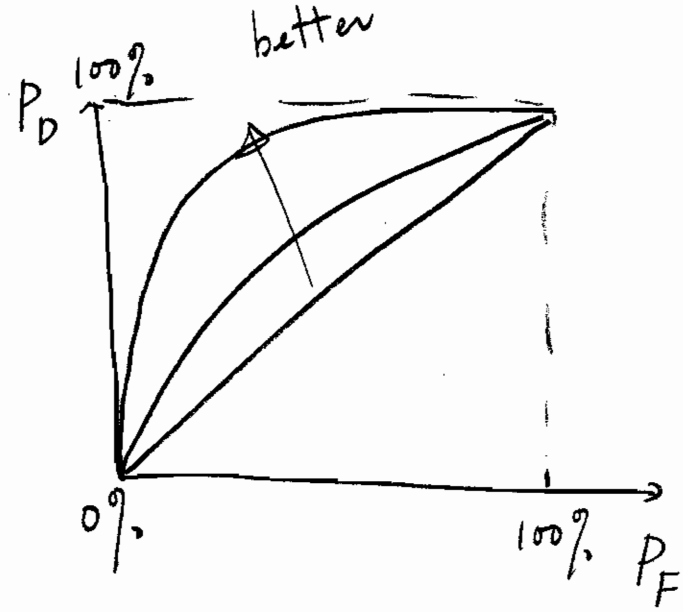
(ii) Neyman-Pearson Criterion

$$P_F = \int_{\lambda}^{\infty} f_{\Lambda|H_0}(\Lambda|H_0)\, d\Lambda = \alpha'$$

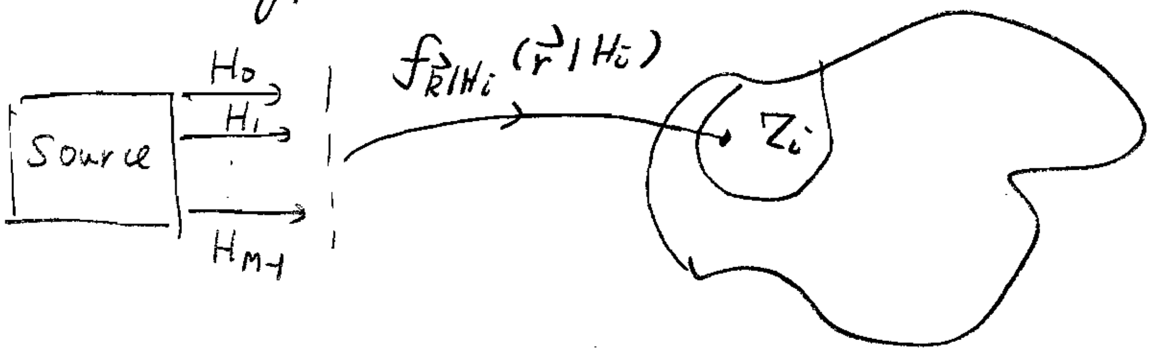
④ $\Lambda(\vec{r})$ can sometimes be reduced to the sufficient statistics $l(\vec{r})$ when the a priori probability density functions are in Gaussian forms.

2.2.2 Receiver Operating Characteristic

We can depict the receiver operating characteristic curve, which are the curves of P_D versus P_F



2.3 M Hypotheses



The M hypotheses are associated with the given M a priori probabilities $P_0, P_1, \cdots, P_{M-1}$.
We can construct the risk or the expected cost as

$$\mathcal{R} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_j C_{ij} \int_{Z_i} f_{\vec{R}|H_j}(\vec{r}\,|\,H_j)\, d\vec{r}$$

If we make any correct decision, the cost C_{ii}, $\forall i$ can be set to zero such that $C_{ii} = 0$, $\forall i$. However, if we make any incorrect decision, we can assign a positive cost C_{ij}, $\forall i \neq j$ such that $C_{ij} > 0$.

Since $f_{\vec{R}|H_j}(\vec{r}|H_j) \geq 0$, we get the optimal Bayes decision rule as

$$P_i \, f_{\vec{R}|H_i}(\vec{r}|H_i) \underset{\text{Choose } H_i}{>} P_j \, f_{\vec{R}|H_j}(\vec{r}|H_j) \quad \forall j \neq i$$

2.4 Estimation Theory

Example: We want to measure a voltage a at a single time instant. The measurement is corrupted by an independent additive zero-mean Gaussian process (random variable n). Thus the observation value r can be formulated as

$$r = a + n.$$

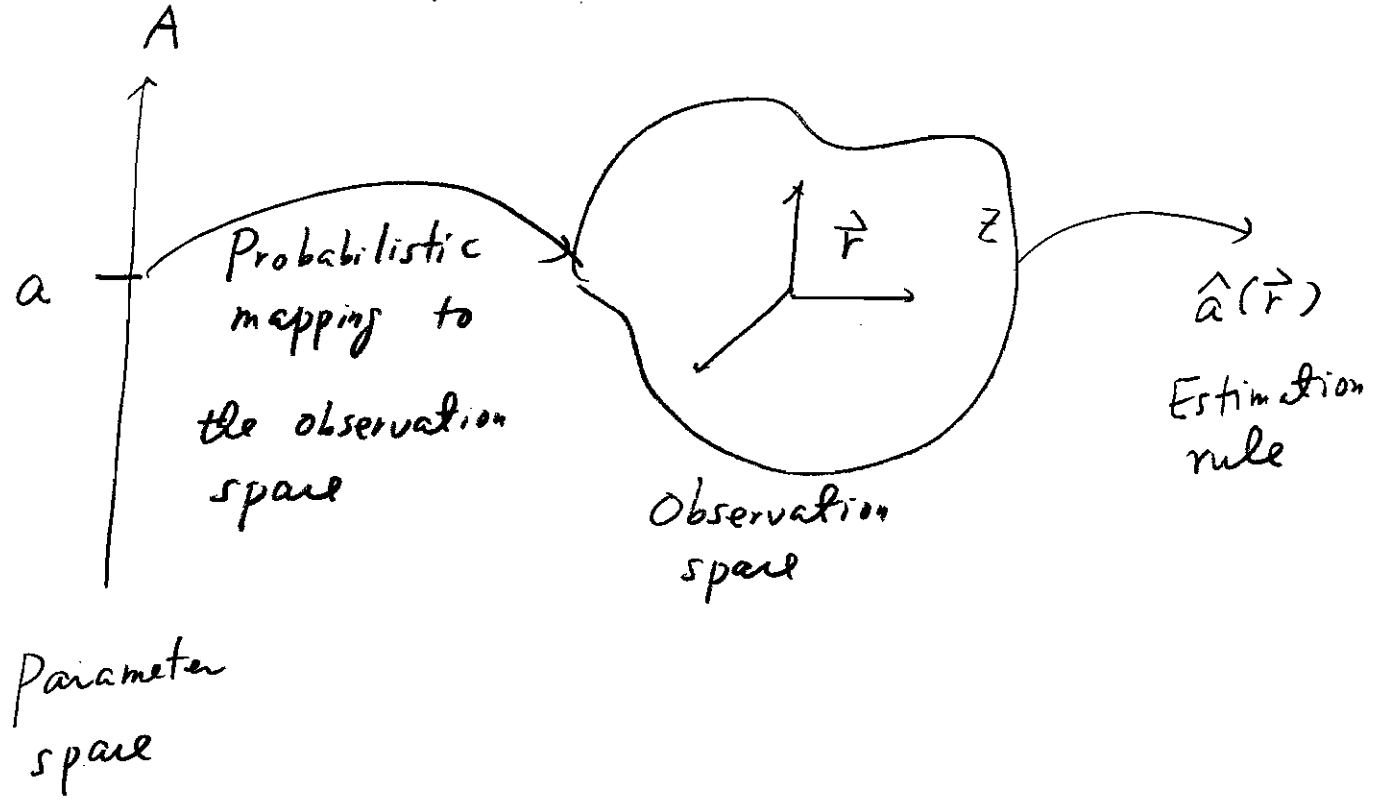
The probability density of the observation can therefore be written as

$$f_{R|a}(r|a) = f_N(r-a)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left\{-\frac{(r-a)^2}{2\sigma_n^2}\right\}$$

The problem is to observe r to estimate a.

This example illustrates the basic features of the estimation problem.



Estimation Model.

Parameter Space: In estimation, the output of the Source is a or multiple parameters (variables). We view this output as a point in a parameter space (A is the example above).

Probabilistic Mapping from the parameter space to the observation space: This is the a priori probability model which relates the variable parameter a to the observation $\vec{r}$.

Observation Space: Usually, it is a finite-dimensional space and we denote a point in it (Z) as the vector $\vec{r}$.

Estimation Rule: After acquiring $\vec{r}$, we shall want to estimate the value of a. The rule is specified as $\hat{a}(\vec{r})$. This function of the observation onto an estimate is the estimation rule.

2.4.1 Random Parameters: Bayes Estimation

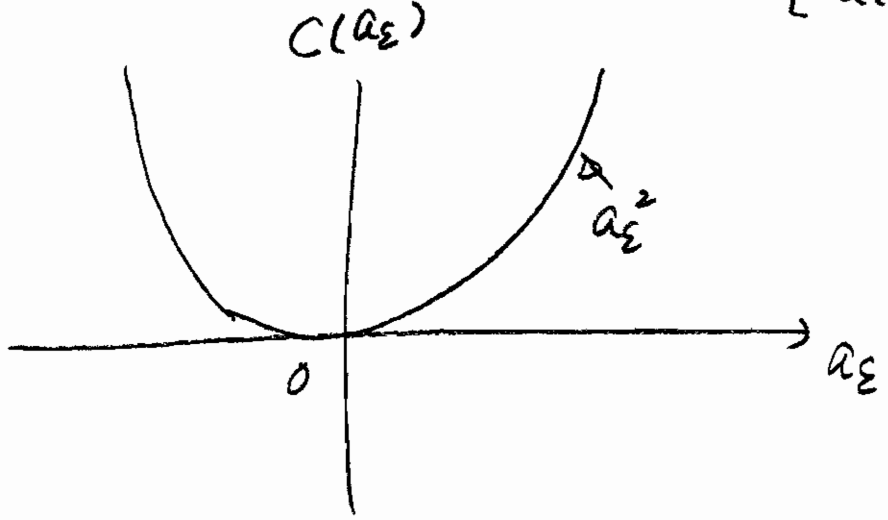
Similar to the Bayes detection criterion, we can construct a cost to the pair of $[a, \hat{a}(\vec{r})]$ over the range of interest in the parameter space. Then an optimization scheme can arise from such a defined cost. In short, we can write such a cost as $C(a, \hat{a})$. A usual cost function is related to the error between the estimate $\hat{a}(\vec{r})$ and the true parameter value a such that

$$a_\varepsilon(\vec{r}) \triangleq \hat{a}(\vec{r}) - a.$$

Based on the above defined error, we can construct three common cost functions for the Bayes estimation

(a) Mean-Square Error : $C(a_\varepsilon) \triangleq a_\varepsilon^2 = a_\varepsilon^2(\vec{r})$

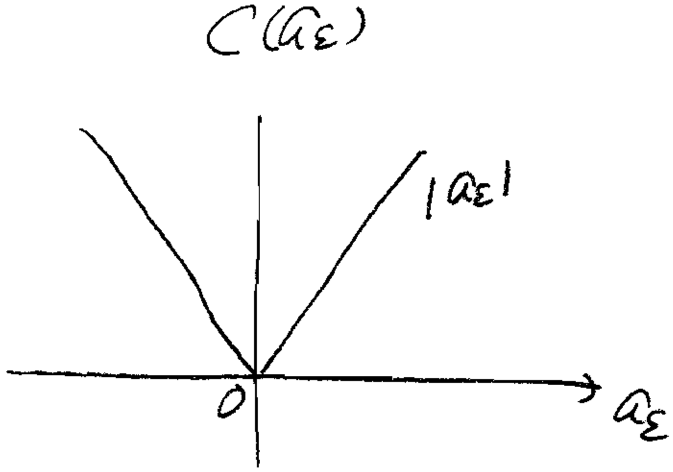
$$= [\hat{a}(\vec{r}) - a]^2$$



It is differentiable!

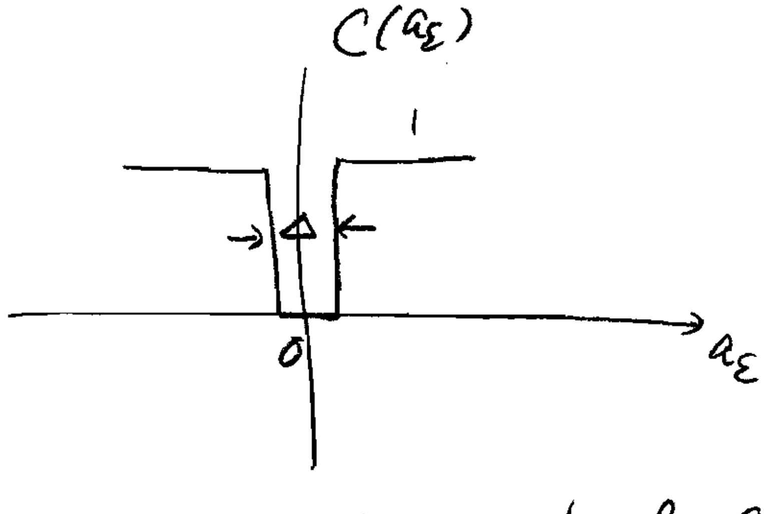
(b) absolute error : $C(a_\varepsilon) \triangleq |a_\varepsilon| = |a_\varepsilon(\vec{r})|$

$= |\hat{a}(\vec{r}) - a|$



It is differentiable except $a_\varepsilon = 0$.

(c) Uniform cost function : $C(a_\varepsilon) \triangleq \begin{cases} 0, & |a_\varepsilon| \le \frac{\Delta}{2} \\ 1, & |a_\varepsilon| > \frac{\Delta}{2} \end{cases}$



It is differentiable except at $a_\varepsilon = \pm \frac{\Delta}{2}$.

Similar to the Bayes detection problem, we can specify a cost function $C[a, \hat{a}(\vec{r})]$ and compute the expected risk as

$$R \triangleq E\{C[a, \hat{a}(\vec{r})]\}$$

$$= \int_{-\infty}^{\infty} da \int_{-\infty}^{\infty} C[a, \hat{a}(\vec{r})] \, f_{a,\vec{R}}(a, \vec{r}) \, d$$

If one of the three error-related cost function is adopted, we have

$$R = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C[a - \hat{a}(\vec{r})] \, f_{a,\vec{R}}(a, \vec{r}) \, d\vec{r} \, da$$

Now, we discuss the estimates based on the three above-mentioned error-related cost functions.

(i) MSE Estimation:

Since $f_{a,\vec{R}}(a, \vec{r}) = f_{\vec{R}}(\vec{r}) \, f_{a|\vec{R}}(a|\vec{r})$,

we can establish the mean-square error criterion as

$$R_{ms} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [a - \hat{a}(\vec{r})]^2 f_{a,\vec{R}}(a,\vec{r})\, d\vec{r}\, da$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [a - \hat{a}(\vec{r})]^2 f_{a|\vec{R}}(a|\vec{r})\, da \times f_{\vec{R}}(\vec{r})\, d\vec{r}$$

The optimal estimation in the mean-square error sense can therefore be achieved when

$$\frac{d R_{ms}}{d\hat{a}} = 0$$

$$\Rightarrow \frac{d}{d\hat{a}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [a - \hat{a}(\vec{r})]^2 f_{a|\vec{R}}(a|\vec{r})\, da \times f_{\vec{R}}(\vec{r})\, d\vec{r} = 0$$

$$\Rightarrow -2\int_{-\infty}^{\infty} a\, f_{a|\vec{R}}(a|\vec{r})\, da + 2\hat{a}(\vec{r})\int_{-\infty}^{\infty} f_{a|\vec{R}}(a|\vec{r})\, da = 0$$

$$\Rightarrow \hat{a}_{ms}(\vec{r}) = \int_{-\infty}^{\infty} a\, f_{a|\vec{R}}(a|\vec{r})\, da = \underbrace{E_A\{a|\vec{r}\}}_{\text{expected mean or conditional mean}}$$

(ii) Mean-Absolute Error Estimation :

$$R_{abs} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\vec{R}}(\vec{r})\, |a - \hat{a}(\vec{r})|\, f_{a|\vec{R}}(a|\vec{r})\, da\, d\vec{r}$$

In order to minimize R_{abs}, we write

$$I(\vec{r}) = \int_{-\infty}^{\hat{a}(\vec{r})} [\hat{a}(\vec{r}) - a]\, f_{a|\vec{R}}(a|\vec{r})\, da + \int_{\hat{a}(\vec{r})}^{\infty} [a - \hat{a}(\vec{r})]\, f_{a|\vec{R}}(a|\vec{r})\, da$$

Then

$$R_{abs} = \int_{-\infty}^{\infty} I(\vec{r})\, d\vec{r}$$

$$\frac{d}{d\hat{a}} R_{abs} = 0$$

$$\Rightarrow \int_{-\infty}^{\hat{a}_{abs}(\vec{r})} f_{a|\vec{R}}(a|\vec{r})\, da = \int_{\hat{a}_{abs}(\vec{r})}^{\infty} f_{a|\vec{R}}(a|\vec{r})\, da$$

$\therefore$ $\hat{a}_{abs}(\vec{r})$ = median of the a posteriori probability density function $f_{a|\vec{R}}(a|\vec{r})$

Example: If the observation r is given by

$$r = a + n,$$

where a is a Gaussian variable with mean m_a. n is the additive Gaussian noise with mean m_n. Determine the $\hat{a}_{abs}(r)$

Solution:

$$f_{A|R}(a|r) = \frac{f_{A,R}(a,r)}{f_R(r)} = \frac{f_{R|A}(r|a)\, f_A(a)}{f_R(r)}$$

$$f_A(a) = \frac{1}{\sqrt{2\pi}\,\sigma_a} \exp\left[-\frac{(a-m_a)^2}{2\sigma_a^2}\right]$$

$$f_{R|A}(r|a) = \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left[-\frac{(r-a)^2}{2\sigma_n^2}\right]$$

$$E[r] = E[a] + E[n] = m_a + m_n$$

$$E[r^2] = E[a^2] + E[n^2] + 2E[a]E[n]$$

$$= \sigma_a^2 + m_a^2 + \sigma_n^2 + m_n^2 + 2 m_a m_n$$

$$\sigma_r^2 = E[r^2] - E^2[r]$$

$$= \sigma_a^2 + \sigma_n^2$$

$$f_{A,R}(a,r) = f_{R|A}(r|a)\, f_A(a)$$

$$f_R(r) = \int_{-\infty}^{\infty} f_{A,R}(a,r)\, da$$

$$f_{A|R}(a|r) = \frac{f_{A,R}(a,r)}{f_R(r)}$$

$$= \frac{1}{2\pi \sigma_n \sigma_n} \exp\left[-\left(\frac{1}{2\sigma_n^2} + \frac{1}{2\sigma_a^2}\right) a^2 + \left(\frac{m_a}{\sigma_a^2} + \frac{r}{\sigma_n^2}\right) a - \left(\frac{m_a^2}{2\sigma_a^2} + \frac{r^2}{2\sigma_n^2}\right)\right] \times \frac{1}{f_R(r)}$$

$$= C \exp\left[-\left(\frac{1}{2\sigma_n^2} + \frac{1}{2\sigma_a^2}\right)(a-d)^2 \right]$$

where C is independent of a and

$$d \triangleq \frac{\frac{m_a}{\sigma_a^2} + \frac{r}{\sigma_n^2}}{\frac{1}{\sigma_n^2} + \frac{1}{\sigma_a^2}}$$

$$= \frac{m_a \sigma_n^2 + r \sigma_a^2}{\sigma_n^2 + \sigma_a^2}$$

is the median of $f_{A|R}(a|r)$

Thus

$$\hat{a}_{abs}(r) = \frac{m_a \sigma_n^2 + r\,\sigma_a^2}{\sigma_a^2 + \sigma_n^2}$$

(iii) Maximum a posteriori estimation:

$$R_{unf} = \int_{-\infty}^{\infty} f_{\vec{R}}(\vec{r}) \left[1 - \int_{\hat{a}_{unf}(\vec{r}) - \frac{\Delta}{2}}^{\hat{a}_{unf}(\vec{r}) + \frac{\Delta}{2}} \right.$$

$$\left. \cdot\ f_{A|\vec{R}}(a|\vec{r})\, da\, d\vec{r} \right.$$

To minimize R_{unf}, we have to maximize

$$\underbrace{\int_{\hat{a}_{unf}(\vec{r}) - \frac{\Delta}{2}}^{\hat{a}_{unf}(\vec{r}) + \frac{\Delta}{2}} f_{A|\vec{R}}(a|\vec{r})\, da}_{\text{the only term in } R_{unf} \text{ associated with } a}$$

Of particular interest, we let $\Delta \to 0$; thus,

$$\int_{\hat{a}_{unf}(\vec{r}) - \frac{\Delta}{2}}^{\hat{a}_{unf}(\vec{r}) + \frac{\Delta}{2}} f_{A|\vec{R}}(a|\vec{r})\, da$$

$$\cong \Delta\, f_{A|\vec{R}}(a|\vec{r})$$

Hence,

$$\arg\max_a \left[\Delta f_{A|\vec{R}}(a|\vec{r}) \right]$$

$$= \arg\max_a \left[f_{A|\vec{R}}(a|\vec{r}) \right]$$

In other words, R_{unf} is minimized when

$f_{A|\vec{R}}(a|\vec{r})$ is at its peak value.

Or equivalently, we would like to find

$\hat{a}(\vec{r})$ such that

$$\left. \frac{\partial \ln \left[f_{A|\vec{R}}(a|\vec{r}) \right]}{\partial a} \right|_{a = \hat{a}(\vec{r})} = 0,$$

Such $\hat{a}(\vec{r}) = \hat{a}_{MAP}(\vec{r})$ is the MAP estimate.

According to Bayes' rule,

$$f_{A|\vec{R}}(a|\vec{r}) = \frac{f_{\vec{R}|A}(\vec{r}|a)\, f_A(a)}{f_{\vec{R}}(\vec{r})}$$

Then,

$$\ln\left[f_{A|\vec{R}}(a|\vec{r})\right] = \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right] + \ln\left[f_A(a)\right] - \ln\left[f_{\vec{R}}(\vec{r})\right]$$

Perserving the two terms on the right handside of the above equation, which are related to a, we get

$$l(a) \triangleq \underbrace{\ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}_{\text{a priori probability}} + \ln\left[f_A(a)\right]$$

The MAP estimate $\hat{a}(\vec{r})$ can also be obtained from the a priori probability such that

$$\left.\frac{\partial\, l(a)}{\partial a}\right|_{a=\hat{a}(\vec{r})} = \left.\frac{\partial \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}{\partial a}\right|_{a=\hat{a}(\vec{r})} + \left.\frac{\partial \ln\left[f_A(a)\right]}{\partial a}\right|_{a=\hat{a}(\vec{r})}$$

$$= 0$$

Example: A series of observations r_i, $i=1,2,\cdots$ can be written as

$$r_i = a + n_i, \quad i = 1, 2, \cdots, N.$$

We assume that a is Gaussian $N(0, \sigma_a)$ and n_i are each

N — normal distribution; 0 — mean; σ_a — variance

independent Gaussian variables $N(0, \sigma_n)$.

Hence,

$$f_{\vec{R}|A}(\vec{r}|a) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left[-\frac{(r_i - a)^2}{2\sigma_n^2}\right]$$

$$f_A(a) = \frac{1}{\sqrt{2\pi}\,\sigma_a} \exp\left[-\frac{a^2}{2\sigma_a^2}\right]$$

First, we would like to find the mean-square error estimate $\hat{a}_{ms}(\vec{r})$ of a. We have

$$f_{A|\vec{R}}(a|\vec{r}) = \left[\frac{\left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_n}\right)\frac{1}{\sqrt{2\pi}\,\sigma_a}}{f_{\vec{R}}(\vec{r})}\right] \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^{N}(r_i - a)^2}{\sigma_n^2} + \frac{a^2}{\sigma_a^2}\right]\right\}$$

$$= k(\vec{r}) \exp\left\{ -\frac{1}{2\sigma_p^2}\left[a - \frac{\sigma_a^2}{\sigma_a^2 + \frac{\sigma_n^2}{N}} \left(\frac{1}{N}\sum_{i=1}^{N} r_i \right)\right]^2 \right\}$$

where $$\sigma_p^2 \triangleq \frac{1}{\frac{1}{\sigma_a^2} + \frac{N}{\sigma_n^2}} = \underbrace{\frac{\sigma_a^2 \sigma_n^2}{N\sigma_a^2 + \sigma_n^2}}_{\text{a posteriori variance}}$$

Since $f_{A|\vec{R}}(a|\vec{r})$ is a Gaussian PDF, the estimate $\hat{a}_{ms}(\vec{r})$ is just the conditional mean or the median value such that

$$\hat{a}_{ms}(\vec{r}) = \frac{\sigma_a^2}{\sigma_a^2 + \frac{\sigma_n^2}{N}} \left(\frac{1}{N}\sum_{i=1}^{N} r_i \right)$$

In this example, we can draw two remarks:

1. The observation r_i enter into the a posteriori density only through their sum, i.e.,

$$l(\vec{r}) = \sum_{i=1}^{N} r_i$$

is a sufficient statistics.

2. The estimation rule is controled by the signal-to-noise ratio σ_a^2 / σ_n^2;

(i) if $\sigma_a^2 \ll \dfrac{\sigma_n^2}{N} \Rightarrow \hat{a}_{ms}(\vec{r}) \to 0$,
low SNR

we can just estimate $\hat{a}_{ms}(\vec{r})$ as the apriori mean $E_A(a) = \int_{-\infty}^{\infty} a\, f_A(a)\, da$,

(ii) if $\sigma_a^2 \gg \dfrac{\sigma_n^2}{N} \Rightarrow \hat{a}_{ms}(\vec{r}) \to \dfrac{1}{N}\sum_{i=1}^{N} r_i$,
high SNR

we can estimate $\hat{a}_{ms}(\vec{r})$ as the arithmetic average of r_i or

$$\lim_{\frac{\sigma_n^2}{N\sigma_a^2} \to 0} \hat{a}_{ms}(\vec{r}) = \frac{1}{N}\sum_{i=1}^{N} r_i .$$

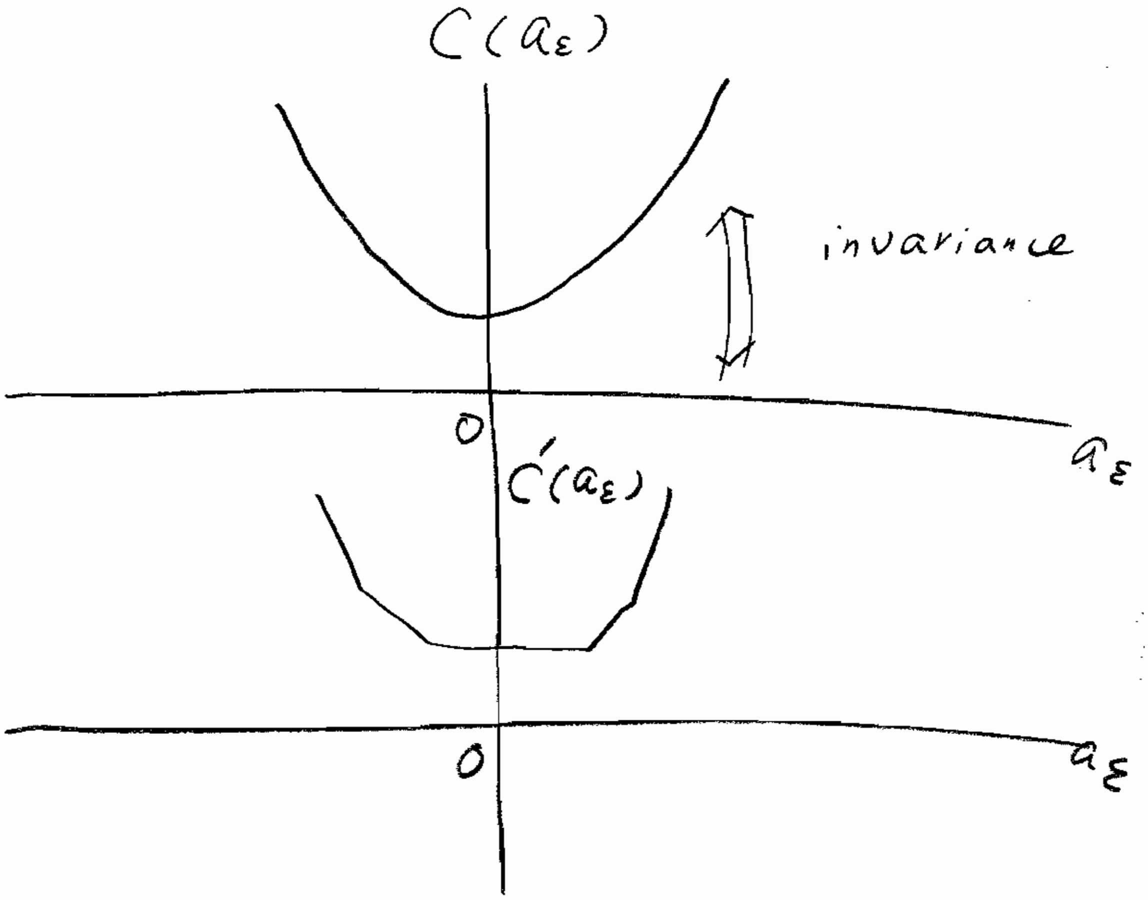
It can be verified that $\hat{a}_{map}(\vec{r}) = \hat{a}_{ms}(\vec{r})$ and $\hat{a}_{abs}(\vec{r}) = \hat{a}_{ms}(\vec{r})$ in this example.

When three estimates $\hat{a}_{ms}(\vec{r})$, $\hat{a}_{map}(\vec{r})$, $\hat{a}_{abs}(\vec{r})$ will lead to the same rule? The invariance to the choice of a cost function among these three in the presence of the two following property 1: We assume that the underlying cost function $C(a_\varepsilon)$ is a symmetric, complex-upward function and that the a posteriori density $f_{A|\vec{R}}(a|\vec{r})$ is symmetric around its conditional mean, i.e.,

$$C(a_\varepsilon) = C(-a_\varepsilon)$$

$$C(b x_1 + (1-b) x_2) \le b\, C(x_1) + (1-b)\, C(x_2)$$

where $0 < b < 1$, $x_1 \neq x_2$



We say the cost function is strictly convex (or upward). Hence, we define

$$z \triangleq a - \hat{a}_{ms} = a - E\{a|\vec{r}\}$$

Proof:

As before, we may minimize the conditional risk as

$$\mathcal{R}_B(\hat{a}|\vec{r}) \triangleq E_A[C(\hat{a}-a)|\vec{r}]$$

$$= E_A[C(a-\hat{a})|\vec{r}].$$

$$\mathcal{R}_B(\hat{a}|\vec{r}) = \int_{-\infty}^{\infty} C(\hat{a}-\hat{a}_{ms}-Z)\, f_{Z|\vec{R}}(z|\vec{r})\,dz$$

$$= \underbrace{\int_{-\infty}^{\infty} C(\hat{a}-\hat{a}_{ms}+Z)\, f_{Z|\vec{R}}(z|\vec{r})\,dz}_{E\{C[z-(\hat{a}_{ms}-\hat{a})]\}}$$

$$\left(f_{Z|\vec{R}}(z|\vec{r}) = f_{Z|\vec{R}}(-z|\vec{r}) \right)$$

$$= \int_{-\infty}^{\infty} C(\hat{a}_{ms}-\hat{a}-z)\, f_{Z|\vec{R}}(z|\vec{r})\,dz$$

$$= \underbrace{\int_{-\infty}^{\infty} C(\hat{a}_{ms}-\hat{a}+z)\, f_{Z|\vec{R}}(z|\vec{r})\,dz}_{E\{C[z+(\hat{a}_{ms}-\hat{a})]\}}$$

Thus,

$$R_B(\hat{a}|\vec{r}) = \frac{1}{2} E\{ C[z + (\hat{a}_{ms} - \hat{a})] | \vec{r} \}$$
$$+ \frac{1}{2} E\{ C[z - (\hat{a}_{ms} - \hat{a})] | \vec{r} \}$$
$$\geq E\{ C[\tfrac{1}{2}(z + (\hat{a}_{ms} - \hat{a}))$$
$$+ \tfrac{1}{2}(z - (\hat{a}_{ms} - \hat{a}))] | \vec{r} \}$$
$$= E\{ C(z) | \vec{r} \}$$

The equality can be achieved if $\hat{a} = \hat{a}_{ms}$.

Consequently the $R_B(\hat{a}|\vec{r})$ is minimized only when $\hat{a} = \hat{a}_{ms}$ (estimation is invariant!)

Property 2: We assume that the cost function is a symmetric, nondecreasing function and that the a posteriori density $f_{A|\vec{R}}(a|\vec{r})$ is a symmetric (about the conditional mean), unimodal function satisfying the condition:

$$\lim_{x \to \infty} C(x) f_{A|\vec{R}}(x|\vec{r}) = 0$$

such that any cost function in this class leads to the corresponding estimate $\hat{a} = \hat{a}_{ms}$.

Proof is neglected.

Example: The variable a appears in the signal in a nonlinear manner. Each observation r_i consists of $\underset{\substack{\text{nonlinear} \\ \text{function}}}{s(a)}$ with a Gaussian random variable n_i, $N(0, \sigma_n)$. n_i are statistically independent of each other and a, such that

$$r_i = s(a) + n_i.$$

Thus, the a posteriori density can be written as

$$f_{A|\vec{R}}(a|\vec{r}) = k(\vec{r}) \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^{N}(r_i - s(a))^2}{\sigma_n^2} + \frac{a^2}{\sigma_a^2}\right]\right\}$$

The MAP estimate can be obtained using

$$\left.\frac{\partial l(a)}{\partial a}\right|_{a=\hat{a}_{MAP}} = \left.\frac{\partial \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}{\partial a}\right|_{a=\hat{a}_{MAP}}$$

$$+ \left. \frac{\partial \ln[f_A(a)]}{\partial a} \right|_{a=\hat{a}_{MAP}} = 0$$

$$\Rightarrow \frac{\partial}{\partial a} \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^{N} (r_i - s(a))^2}{\sigma_n^2} + \frac{a^2}{\sigma_a^2} \right] - \ln[f_A(a)] + \ln[f_A(a)] \right\} = 0$$

$$\Rightarrow -\frac{1}{\sigma_n^2} \sum_{i=1}^{N} (r_i - s(a)) \frac{\partial s(a)}{\partial a} + \frac{a}{\sigma_a^2} = 0$$

$$\Rightarrow \hat{a}_{MAP}(\vec{r}) = \frac{\sigma_a^2}{\sigma_n^2} \sum_{i=1}^{N} (r_i - s(a)) \left. \frac{\partial s(a)}{\partial a} \right|_{a=\hat{a}_{MAP}(\vec{r})}$$

Concluding Remarks:

1. The minimum mean-square error estimate (MMSE) is always the mean of the a posteriori density (the conditional mean).

2. The maximum a posteriori estimate (MAP) is the value of a at which the a posteriori density has its maximum.

3. For a large class of cost functions the optimum estimate is the conditional mean whenever the a posteriori density is a unimodal function which is symmetric about the conditional mean

2.4.2 Real (Nonrandom Parameter) Estimation

If the unknown parameter is assumed to be nonrandom, we need to modify the previously discussed Bayes procedure to eliminate the average over $f_A(a)$. For example, consider a mean-square error criterion such that

$$R(a) \triangleq \int_{-\infty}^{\infty} [\hat{a}(\vec{r}) - a]^2 f_{\vec{R}|A}(\vec{r}|a)\, d\vec{r}$$

where the expectation is only over $\vec{r}$, for it is the only random variable. Minimizing $R(a)$, we obtain

$$\hat{a}_{ms}(\vec{r}) = a.$$

However, $\hat{a}_{ms}(\vec{r}) = a$ totally depends on the unknown a, which is not realistic in practice. Thus, we need to measure the quality of the estimates $\hat{a}(\vec{r})$. The first measure of quality to be considered is the expectation of the estimate:

$$E\{\hat{a}(\vec{r})\} \triangleq \int_{-\infty}^{\infty} \hat{a}(\vec{r}) f_{\vec{R}|A}(\vec{r}|a)\, d\vec{r}$$

The possible values of the expectation can be categorized into the three following cases:

1. If $E\{\hat{a}(\vec{r})\} = a$, for all values of a, we say that the estimate is unbiased.

2. If $E\{\hat{a}(\vec{r})\} = a + b$, where b is not a function of a, we say that the estimate has a known bias.

3. If $E\{\hat{a}(\vec{r})\} = a + b(a)$, we say that the estimate has an unknown bias.

A second measure of quality is the variance of estimation error:

$$Var\,[\hat{a}(\vec{r}) - a] = E\{[\hat{a}(\vec{r}) - a]^2\} - b^2(a).$$

This provides a measure of the spread of the error. In general, there is no straight-forward minimization procedure that will lead to the minimum variance unbiased estimate.

-X. Maximum Likelihood Estimation

Consider a simple estimation problem that the observation can be modeled as

$$r = a + n,$$

where a is an unknown nonrandom parameter and n is the Gaussian noise, $N(0, \sigma_n)$.

The a priori PDF is written as

$$f_{\vec{R}|A}(\vec{r}|a) = \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left[-\frac{1}{2\sigma_n^2}(r-a)\right.$$

The maximum likelihood estimation (MLE) is to estimate the value of a that most likely causes a given value of r to occur. In this simple example, $f_{R|A}(r|a)$ reaches its peak (most probable value) when

$$\hat{a}_{ml}(r) = r.$$

Following this simple example, we can construct the likelihood function $f_{\vec{R}|A}(\vec{r}|a)$ for the nonrandom parameter estimation, such that the maximum likelihood estimate $\hat{a}_{ml}(\vec{r})$ is the value of a at which the likelihood function reaches its maximum.

If such a maximum value is interior to the range of a and $\ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]$ has a continuous first derivative, then a necessary condition on $\hat{a}_{ml}(\vec{r})$ is obtained by differentiating the log-likelihood function $\ln[f_{\vec{R}|A}(\vec{r}|a)]$ with respect to a and setting the result equal to zero:

$$\left. \frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a} \right|_{a=\hat{a}_{ml}(\vec{r})} = 0$$

The above equation is called the likelihood equation. Comparing this equation with Eq. (137) in Van Trees' text, we can see that

$$\left. \frac{\partial l(a)}{\partial a} \right|_{a=\hat{a}_{MAP}} = \left. \frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a} \right|_{a=\hat{a}_{ml}} = 0$$

when $\frac{\partial \ln[f_A(a)]}{\partial a} = 0$ is assumed

X. Cramer-Rao Inequality (lower bound) or CRLB

For the same given set of observations, we may formulate many estimation rules (estimators). However, according to the previous discussion, we have to "measure" the quality of any estimator. The variance is a popular measure of the estimates. To provide the variance of any estimate $\hat{a}(\vec{r})$ of the real value a, we shall prove the following theorem

Theorem: If $\hat{a}(\vec{r})$ is any unbiased estimate of a, then

(i) $$\mathrm{Var}\left[\hat{a}(\vec{r}) - a\right] \geq \frac{1}{E\left\{\left(\frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a}\right)^2\right\}}$$

Or equivalently,

(ii)

$$\mathrm{var}\left[\hat{a}(\vec{r})-a\right] \geq \frac{-1}{E\left\{\dfrac{\partial^2 \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}{\partial a^2}\right\}}$$

where

$$\frac{\partial f_{\vec{R}|A}(\vec{r}|a)}{\partial a} \quad \text{and} \quad \frac{\partial^2 f_{\vec{R}|A}(\vec{r}|a)}{\partial a^2}$$

both exist and are absolutely integrable

Proof:

Since $\hat{a}(\vec{r})$ is unbiased, we get

$$E\left\{\hat{a}(\vec{r})-a\right\} = \int_{-\infty}^{\infty}\left[\hat{a}(\vec{r})-a\right] f_{\vec{R}|A}(\vec{r}|a)\, d\vec{r}$$

$$= 0$$

$$\frac{d}{da}\int_{-\infty}^{\infty} f_{\vec{R}|A}(\vec{r}|a)\left[\hat{a}(\vec{r})-a\right] d\vec{r}$$

$$= \int_{-\infty}^{\infty} \frac{\partial}{\partial a}\left\{ f_{\vec{R}|A}(\vec{r}|a)\left[\hat{a}(\vec{r})-a\right] d\vec{r} = 0.\right.$$

differentiable

Then

$$-\underbrace{\int_{-\infty}^{\infty} f_{\vec{R}|A}(\vec{r}|a)\,d\vec{r}}_{\substack{\shortparallel \\ 1}} + \int_{-\infty}^{\infty} \frac{\partial f_{\vec{R}|A}(\vec{r}|a)}{\partial a}\left[\hat{a}(\vec{r}) - a\right] d\vec{r} = 0$$

Observe that

$$\frac{\partial f_{\vec{R}|A}(\vec{r}|a)}{\partial a} = \frac{\partial \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}{\partial a} f_{\vec{R}|A}(\vec{r}|a)$$

Thus, we have

$$\int_{-\infty}^{\infty} \frac{\partial \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}{\partial a} \underbrace{f_{\vec{R}|A}(\vec{r}|a)}_{\sqrt{f_{\vec{R}|A}(\vec{r}|a)}\,\sqrt{f_{\vec{R}|A}(\vec{r}|a)}} \left[\hat{a}(\vec{r}) - a\right] d\vec{r} = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} \left[\frac{\partial \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}{\partial a}\sqrt{f_{\vec{R}|A}(\vec{r}|a)}\right]\left[\sqrt{f_{\vec{R}|A}(\vec{r}|a)}\,\left[\hat{a}(\vec{r}) - a\right]\right] d\vec{r} = 1$$

Using the Schwartz Inequality, we get

$$\left\{\int_{-\infty}^{\infty}\left[\frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a}\right]^2 f_{\vec{R}|A}(\vec{r}|a)\,d\vec{r}\right\}\left\{\int_{-\infty}^{\infty}[\hat{a}(\vec{r})-a]^2 f_{\vec{R}|A}(\vec{r}|a)\,d\vec{r}\right\}\geq$$

where the equality holds if and only if

$$\frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a} = [\hat{a}(\vec{r})-a]\,k(a),$$

$\forall \vec{r}$ and $\forall a$. Consequently,

$$\mathrm{Var}[\hat{a}(\vec{r})-a] = E\{[\hat{a}(\vec{r})-a]^2\}$$

$$\geq \frac{1}{E\left\{\left(\frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a}\right)^2\right\}}$$

On the other hand,

$$\int_{-\infty}^{\infty} f_{\vec{R}|A}(\vec{r}|a)\,d\vec{r} = 1$$

Differentiating it with respect to a, we get

$$\int_{-\infty}^{\infty} \frac{\partial f_{\vec{R}|A}(\vec{r}|a)}{\partial a} d\vec{r} = \int_{-\infty}^{\infty} \frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a} \times f_{\vec{R}|A}(\vec{r}|a)\, d\vec{r} = 0$$

In addition, differentiating again with respect to a and employing Eq. (183), we have

$$\int_{-\infty}^{\infty} \frac{\partial^2 \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a^2} f_{\vec{R}|A}(\vec{r}|a)\, d\vec{r} + \int_{-\infty}^{\infty} \left(\frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a}\right)^2 f_{\vec{R}|A}(\vec{r}|a)\, d\vec{r} = 0$$

Or

$$E\left\{\frac{\partial^2 \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a^2}\right\} = -E\left\{\frac{\partial \ln[f_{\vec{R}|A}(\vec{r}|a)]}{\partial a}\right\}$$

Concluding Remarks on CRLB:

1. It shows that any unbiased estimate $\hat{a}(\vec{r})$ must have a variance greater than the CRLB.
2. If Eq. (187) holds, the estimate $\hat{a}_{ml}(\vec{r})$ will satisfy the bound with an equality. We can

show this by combining Eqs. (187) and (177).

According to Eq (187), we have

$$0 = \frac{\partial \ln [f_{\vec{R}|A} (\vec{r}|a)]}{\partial a} \Big|_{a = \hat{a}_{ml}(\vec{r})}$$

$$= [\hat{a}(\vec{r}) - a] \, k(a) \Big|_{a = \hat{a}_{ml}(\vec{r})}$$

$$\Rightarrow \hat{a}(\vec{r}) = \hat{a}_{ml}(\vec{r}) \quad \text{or} \quad k(\hat{a}_{ml}) = 0.$$

It means that a unique solution to the likelihood equation should be obtained if an efficient estimate (minimum-variance estimate) exists.

3. If an efficient estimate (minimum-variance estimate) does not exist, i.e.,

$$\frac{\partial \ln [f_{\vec{R}|A} (\vec{r}|a)]}{\partial a} \neq [\hat{a}(\vec{r}) - a] \, k(a),$$

we do not know how good $\hat{a}_{ml}(\vec{r})$ is. We do not know how close any estimate will be to the CRLB.

4. In order to use the CRLB, we must verify the estimate is unbiased.

Example: Assume a is a nonrandom parameter and n_i are i.i.d. Gaussian noise, $i=1,2,\ldots,N$. We have the observations r_i, $i=1,2,\ldots,N$, such that

$$r_i = a + n_i, \quad i=1,2,\ldots,N.$$

Thus,

$$f_{\vec{R}|A}(\vec{r}|a) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left[-\frac{(r_i-a)^2}{2\sigma_n^2}\right].$$

$$\frac{\partial \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}{\partial a} = \frac{N}{\sigma_n^2}\left(\frac{1}{N}\sum_{i=1}^{N} r_i - a\right).$$

Hence, $\hat{a}_{ml}(\vec{r}) = \frac{1}{N}\sum_{i=1}^{N} r_i.$

$$E\{\hat{a}_{ml}(\vec{r})\} = E\left\{\frac{1}{N}\sum_{i=1}^{N} r_i\right\}$$

$$= \frac{1}{N}\sum_{i=1}^{N} E\{r_i\}$$

$$= \frac{1}{N}\sum_{i=1}^{N} a = a$$

Consequently, $\hat{a}_{ml}(\vec{r})$ is unbiased.

To evaluate the variance of the estimate, we calculate

$$\frac{\partial^2 \ln [f_{\vec{R}|A}(\vec{r}|a)]}{\partial a^2} = -\frac{N}{\sigma_n^2}.$$

According to Eq. (179), we have

$$\mathrm{var}[\hat{a}_{ml}(\vec{r}) - a] = \frac{\sigma_n^2}{N}$$

On the other hand, we may verify

$$\begin{aligned}
&\mathrm{var}[\hat{a}_{ml}(\vec{r}) - a] \\
&= E\{[\hat{a}_{ml}(\vec{r}) - a]^2\} \\
&= E\{[\hat{a}_{ml}(\vec{r})]^2\} - a^2 \\
&= E\{[\frac{1}{N}\sum_{i=1}^{N} r_i]^2\} - a^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\{r_i r_j\} - a^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\{(a+n_i)(a+n_j)\} - a^2 \\
&= \frac{1}{N^2}(N^2 a^2 + N E\{n_i^2\}) - a^2 \\
&= a^2 + \frac{\sigma_n^2}{N} - a^2 = \frac{\sigma_n^2}{N}
\end{aligned}$$

Hence $\hat{a}_{me}(\vec{r})$ achieves the CRLB!

Theorem: Let a be a random variable and $\vec{r}$, the observation vector. The mean-square error of any estimate $\hat{a}(\vec{r})$ satisfies the following inequality:

$$E\left\{\left[\hat{a}(\vec{r})-a\right]^2\right\} \geq \frac{1}{E\left\{\left[\frac{\partial \ln\left[f_{\vec{R},A}(\vec{r},a)\right]}{\partial a}\right]^2\right\}}$$

$$= \frac{1}{-E\left\{\frac{\partial^2 \ln\left[f_{\vec{R},A}(\vec{r},a)\right]}{\partial a^2}\right\}}$$

Observe that the probability density function involved above is the joint density instead of the conditional density in the previous discussion in the real nonrandom estimation. Hence, the expectation here is over both a and $\vec{r}$. Moreover, the three following conditions are assumed to exist:

1. $\dfrac{\partial f_{\vec{R},A}(\vec{r}, a)}{\partial a}$ is absolutely integrable with respect to $\vec{r}$ and a.

2. $\dfrac{\partial^2 f_{\vec{R},A}(\vec{r}, a)}{\partial a^2}$ is absolutely integrable with respect to $\vec{r}$ and a.

3. The conditional expectation of the estimation error given a is

$$B(a) = \int_{-\infty}^{\infty} [\hat{a}(\vec{r}) - a] \, f_{\vec{R}|A}(\vec{r}\,|\,a) \, d\vec{r}$$

We assume that

$$\lim_{a \to \infty} B(a) f_A(a) = 0$$

and $\displaystyle \lim_{a \to -\infty} B(a) f_A(a) = 0$

Proof:

$$\frac{d}{da}\left[f_A(a)\,B(a)\right] = -\int_{-\infty}^{\infty} f_{\vec{R},A}(\vec{r},a)\,d\vec{r} + \int_{-\infty}^{\infty} \frac{\partial f_{\vec{R},A}(\vec{r},a)}{\partial a}\left[\hat{a}(\vec{r})-a\right] d\vec{r}$$

$$\int_{-\infty}^{\infty} \frac{d}{da}\left[f_A(a)\,B(a)\right] da = f_A(a)\,B(a)\Big|_{-\infty}^{\infty}$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\vec{R},A}(\vec{r},a)\,d\vec{r}\,da + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial f_{\vec{R},A}(\vec{r},a)}{\partial a}\left[\hat{a}(\vec{r})-a\right] d\vec{r}\,da$$

$$= -1 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial f_{\vec{R},A}(\vec{r},a)}{\partial a}\left[\hat{a}(\vec{r})-a\right] da\,d\vec{r}$$

$$\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial f_{\vec{R},A}(\vec{r},a)}{\partial a}\left[\hat{a}(\vec{r})-a\right] da\,d\vec{r} = 1$$

Similar to Eqs (182)-(188) in Van Trees' text, we get

$$E\left\{\left[\hat{a}(\vec{r})-a\right]^2\right\} \geq \frac{1}{E\left\{\left(\frac{\partial \ln\left[f_{\vec{R},A}(\vec{r},a)\right]}{\partial a}\right)^2\right\}}$$

or equivalently,

$$E\left\{\left[\hat{a}(\vec{r})-a\right]^2\right\} \geq \frac{1}{-E\left\{\frac{\partial^2 \ln\left[f_{\vec{R}|A}(\vec{r}|a)\right]}{\partial a^2}\right\} - E\left\{\frac{\partial^2 \ln\left[f_A(a)\right]}{\partial a^2}\right\}}$$

with the equality if and only if

$$\frac{\partial \ln [f_{\vec{R},A}(\vec{r},a)]}{\partial a} = k[\hat{a}(\vec{r}) - a],$$

$\forall \vec{r}$ and $\forall a$.

Hence, $\dfrac{\partial^2 \ln [f_{\vec{R},A}(\vec{r},a)]}{\partial a^2} = -k$

and $\dfrac{\partial^2 \ln [f_{A|\vec{R}}(a|\vec{r})]}{\partial a^2} = -k$.

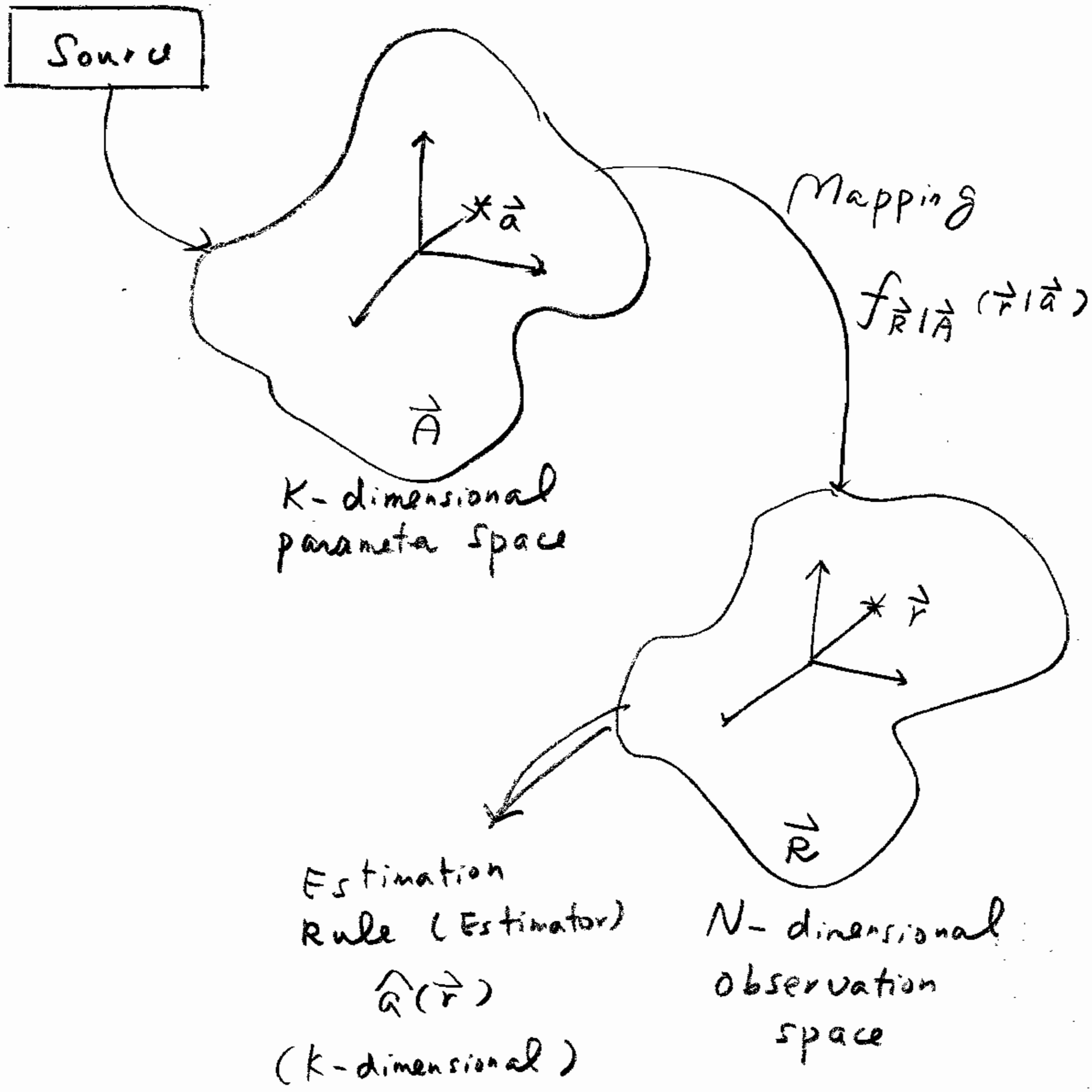
Integrating (227) twice, we have

$$f_{A|\vec{R}}(a|\vec{r}) = \exp[-ka^2 + C_1 a + C_2]$$

and Eq. (228) is simply a statement that the a posteriori probability density of a must be Gaussian for all $\vec{r}$ in order for the minimum-variance estimation. Also, Eq. (226) shows that the MAP estimate will be efficient. It tells us that $\hat{a}_{ms}(\vec{r}) = \hat{a}_{MAP}(\vec{r})$ whenever an efficient estimate exists.

2.4.3 Multiple Parameter Estimation

Most of the ideas and techniques can be extended from 2.4.2 to this section. The model for the multiple parameter estimation is shown as below:



Three issues essential to the multiple parameter estimation is discussed as follows:

(1.) Estimation Procedure

For random variables $\vec{a}$, we could consider the general case of Bayes estimation in which we minimize the risk for some arbitrary scalar cost function $C(\vec{a}, \hat{\vec{a}})$ related to the error vector $\hat{\vec{a}}_\varepsilon(\vec{r})$ such that

$$\vec{a}_\varepsilon(\vec{r}) = \begin{bmatrix} \hat{a}_1(\vec{r}) - a_1 \\ \hat{a}_2(\vec{r}) - a_2 \\ \vdots \\ \hat{a}_K(\vec{r}) - a_K \end{bmatrix} = \hat{\vec{a}}(\vec{r}) - \underset{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix}}{\underset{\parallel}{\vec{a}}}$$

$$= \Big[a_{\varepsilon_i}(\vec{r}) \Big]_{i=1,2,\cdots,K}$$

For a mean-square error criterion, the cost function is simply

$$C(\vec{a}_\varepsilon(\vec{r})) \triangleq \sum_{i=1}^{K} a_{\varepsilon_i}^2(\vec{r}) = \vec{a}_\varepsilon^T(\vec{r})\, \vec{a}_\varepsilon(\vec{r})$$

The expected risk is therefore.

$$R_{ms} = \iint_{-\infty}^{\infty} C(\vec{a}_\varepsilon(\vec{r}))\, f_{\vec{R},\vec{A}}(\vec{r},\vec{a})\, d\vec{r}\, d\vec{a}$$

Or
$$R_{ms} = \int_{-\infty}^{\infty} f_{\vec{R}}(\vec{r}) \times \int_{-\infty}^{\infty} \left[\sum_{i=1}^{k} (\hat{a}_i(\vec{r}) - a_i)^2 \right] \times f_{\vec{A}|\vec{R}}(\vec{a}|\vec{r})\, d\vec{a}\, . d\vec{r}.$$

Hence,
$$\hat{a}_{ms_i}(\vec{r}) = \int_{-\infty}^{\infty} a_i\, f_{\vec{A}|\vec{R}}(\vec{a}|\vec{r})\, d\vec{a}, \quad i = 1, 2, \cdots, k$$

or
$$\hat{\vec{a}}_{ms}(\vec{r}) = \int_{-\infty}^{\infty} \vec{a}\, f_{\vec{A}|\vec{R}}(\vec{a}|\vec{r})\, d\vec{a}$$

The mean-square estimation is invariant over any linear transformation, i.e., if $\vec{b} = \tilde{D}\vec{a}$, where $\tilde{D} \in \mathbb{R}^{L \times K}$, the risk function can be formulated as

$$R'_{ms} = E\left\{ \vec{b}_\varepsilon^{\,T}(\vec{r})\, \vec{b}_\varepsilon(\vec{r}) \right\} = E\left\{ \sum_{i=1}^{L} b_{\varepsilon_i}^2(\vec{r}) \right\}$$

$$\vec{b}_\varepsilon(\vec{r}) = \hat{\vec{b}}(\vec{r}) - \vec{b}.$$

Consequently,

$$\hat{\vec{b}}_{ms}(\vec{r}) = \tilde{D}\,\hat{\vec{a}}_{ms}(\vec{r})$$

On the other hand, for the MAP estimation, we can maximize $f_{\vec{A}|\vec{R}}(\vec{a}|\vec{r})$ to determine $\hat{\vec{a}}(\vec{r})$. If the maximum of $\frac{\partial}{\partial \vec{a}} f_{\vec{A}|\vec{R}}(\vec{a}|\vec{r})$ exists and it is the interior, we can form the set of K simultaneous equations:

$$\left.\frac{\partial \ln f_{\vec{A}|\vec{R}}(\vec{a}|\vec{r})}{\partial a_i}\right|_{\vec{a} = \hat{\vec{a}}_{MAP}(\vec{r})} = 0,$$

$$i = 1, 2, \ldots, K.$$

Accordingly, we can define a partial derivative matrix operator such that

$$\vec{\nabla}_{\vec{a}} \triangleq \begin{bmatrix} \frac{\partial}{\partial a_1} \\ \frac{\partial}{\partial a_2} \\ \vdots \\ \frac{\partial}{\partial a_K} \end{bmatrix}$$

NOT the gradient operator

This operator can be applied only to $1 \times m$ vectors, for example,

$$\vec{\nabla}_{\vec{a}} \vec{G} = \begin{bmatrix} \frac{\partial G_1}{\partial a_1} & \frac{\partial G_2}{\partial a_1} & \cdots & \frac{\partial G_m}{\partial a_1} \\ \vdots & \vdots & \ddots & \\ \frac{\partial G_1}{\partial a_k} & \frac{\partial G_2}{\partial a_k} & \cdots & \frac{\partial G_m}{\partial a_k} \end{bmatrix}$$

Properties of $\vec{\nabla}_{\vec{a}}$:

1. The vector $\vec{a}$ is $n \times 1$ and the vector $\vec{b}$ is $n \times 1$. Then,

$$\vec{\nabla}_{\vec{x}} (\vec{a}^T \vec{b}) = (\vec{\nabla}_{\vec{x}} \vec{a}^T) \vec{b} + (\vec{\nabla}_{\vec{x}} \vec{b}^T) \vec{a}$$

Proof: Since $\vec{a}^T \vec{b}$ is 1×1 $(m=1)$, we have $\vec{c} = \vec{\nabla}_{\vec{x}} (\vec{a}^T \vec{b})$ is $k \times 1$ (assume that $\vec{x}$ is $k \times 1$).

$$\vec{a}^T \vec{b} = \sum_{i=1}^{n} a_i b_i$$

$$\vec{c} = [c_\ell]_{\ell=1,2,\ldots,k}$$

$$c_\ell = \frac{\partial}{\partial x_\ell}(\vec{a}^T\vec{b})$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial x_\ell}(a_i b_i)$$

$$= \sum_{i=1}^{n} \underbrace{\frac{\partial a_i}{\partial x_\ell}}_{(\vec{\nabla}_{\vec{x}}\vec{a}^T)_{\ell,i}} b_i + \sum_{i=1}^{n} a_i \underbrace{\frac{\partial b_i}{\partial x_\ell}}_{(\vec{\nabla}_{\vec{x}}\vec{b}^T)_{\ell,i}}$$

$$= (\vec{\nabla}_{\vec{x}}\vec{a}^T)\,\vec{b} + (\vec{\nabla}_{\vec{x}}\vec{b}^T)\,\vec{a}$$

\#

2. If the $n \times 1$ vector $\vec{b}$ is not a function of $\vec{x}$, then

$$\vec{\nabla}_{\vec{x}}(\vec{b}^T\vec{x}) = \vec{b}$$

Proof: According to Property 1; we get

$$\vec{\nabla}_{\vec{x}}(\vec{b}^T\vec{x}) = \underbrace{(\vec{\nabla}_{\vec{x}}\vec{b}^T)}_{\tilde{0}}\vec{x} + \underbrace{(\vec{\nabla}_{\vec{x}}\vec{x}^T)}_{\tilde{I}}\vec{b}$$

($k \times k$ identity matrix)

$$= \vec{b}$$ \#

3. Let $\vec{c}$ be an $n \times 1$ constant vector, we have

$$\vec{\nabla}_{\vec{x}} (\vec{x}^T \vec{c}) = \vec{c}.$$

Proof: Similar to the property 2, we have

$$\vec{\nabla}_{\vec{x}} (\vec{x}^T \vec{c}) = (\underbrace{\vec{\nabla}_{\vec{x}} \vec{x}^T}_{= \tilde{I}}) \vec{c} + (\underbrace{\vec{\nabla}_{\vec{x}} \vec{c}^T}_{= \tilde{0}}) \vec{x}$$

$$= \vec{c} \quad \#$$

4. $\vec{\nabla}_{\vec{x}} (\vec{x}^T) = \tilde{I}$ ($k \times k$ identity matrix).

$$\vec{\nabla}_{\vec{x}} = \left[\frac{\partial x_i}{\partial x_j} \right]_{1 \le i, j \le k}$$

$$= \tilde{I}_{k \times k} \quad \#$$

5. If a scalar $Q = \vec{a}^T(\vec{x}) \tilde{\Lambda} \vec{a}(\vec{x})$, where $\tilde{\Lambda}$ is a symmetric nonnegative definite $m \times m$ matrix, $\vec{a}(\vec{x})$ is an $m \times 1$ vector. Then

$$\vec{\nabla}_{\vec{x}} Q = 2 (\vec{\nabla}_{\vec{x}} \vec{a}^T(\vec{x})) \tilde{\Lambda} \vec{a}(\vec{x}).$$

Proof: According to Property 1, we have

$$\vec{\nabla}_{\vec{x}} Q = \left(\vec{\nabla}_{\vec{x}} \vec{a}^T(\vec{x}) \right) \tilde{\Lambda} \vec{a}(\vec{x}) + \left[\vec{\nabla}_{\vec{x}} \left(\vec{a}^T(\vec{x}) \tilde{\Lambda} \right) \right] \vec{a}(\vec{x})$$

Since $\vec{\nabla}_{\vec{x}} \left(\vec{a}^T(\vec{x}) \tilde{\Lambda} \right)$

$$= \left(\vec{\nabla}_{\vec{x}} \vec{a}^T(\vec{x}) \right) \tilde{\Lambda} + \underbrace{\left(\vec{\nabla}_{\vec{x}} \tilde{\Lambda} \right)}_{= \tilde{0}} \vec{a}(\vec{x})$$

$$= \left(\vec{\nabla}_{\vec{x}} \vec{a}^T(\vec{x}) \right) \tilde{\Lambda},$$

we have

$$\vec{\nabla}_{\vec{x}} Q = \left(\vec{\nabla}_{\vec{x}} \vec{a}^T(\vec{x}) \right) \tilde{\Lambda} \vec{a}(\vec{x}) + \left(\vec{\nabla}_{\vec{x}} \vec{a}^T(\vec{x}) \right) \tilde{\Lambda} \vec{a}(\vec{x})$$

$$= 2 \left(\vec{\nabla}_{\vec{x}} \vec{a}^T(\vec{x}) \right) \tilde{\Lambda} \vec{a}(\vec{x}) \quad \#$$

6. If $\underbrace{\vec{a}(\vec{x})}_{m \times 1} = \underbrace{\tilde{B}}_{m \times k} \underbrace{\vec{x}}_{k \times 1}$, $Q = \vec{a}^T(\vec{x}) \tilde{\Lambda} \vec{a}(\vec{x})$,

then

$$\vec{\nabla}_{\vec{x}} Q = 2 \tilde{B}^T \tilde{\Lambda} \tilde{B} \vec{x}.$$

Proof:

$$\vec{\nabla}_{\vec{x}}\, \vec{a}^T(\vec{x}) = \vec{\nabla}_{\vec{x}}\, (\tilde{B}\vec{x})^T$$

$$= \vec{\nabla}_{\vec{x}}\, (\vec{x}^T \tilde{B}^T)$$

$$\triangleq \tilde{C}$$

$$(\vec{x}^T\tilde{B}^T)_j = \sum_{l=1}^{K} x_l B_{jl} \quad , \text{ where } \tilde{B} = [B_{lj}]$$

$$\tilde{C} = [C_{ij}] = \frac{\partial}{\partial x_i}(\vec{x}^T\tilde{B}^T)_j$$

$$= \sum_{l=1}^{K} \frac{\partial x_l}{\partial x_i} B_{jl}$$

$$= B_{ji}$$

$$\therefore \tilde{C} = \vec{\nabla}_{\vec{x}}\, \vec{a}^T(\vec{x}) = \tilde{B}^T$$

Since according to Property 5,

$$\vec{\nabla}_{\vec{x}} Q = 2\left(\underbrace{\vec{\nabla}_{\vec{x}}\, \vec{a}^T(\vec{x})}\right) \tilde{\Lambda}\, \vec{a}(\vec{x})$$

$$= 2\tilde{B}^T\tilde{\Lambda}\tilde{B}\vec{x} \quad \#$$

7. If $Q = \vec{x}^T\tilde{\Lambda}\vec{x}$, then $\vec{\nabla}_{\vec{x}} Q = 2\tilde{\Lambda}\vec{x}$.

Proof: According to Property 6, set $\tilde{B} = \tilde{I}$, and we get $\vec{\nabla}_{\vec{x}} Q = 2\tilde{\Lambda}\vec{x}$. #

According to the properties, Eq. (258) can be written in the matrix form such that

$$\overrightarrow{\nabla_{\vec{a}}} \left[\ln \left(f_{\vec{A}|\vec{R}}(\vec{a}|\vec{r}) \right) \right] \Big|_{\vec{a} = \hat{\vec{a}}_{MAP}(\vec{r})} = \vec{0}$$

On the other hand, for the ML estimates, we can form the likelihood equations as

$$\overrightarrow{\nabla_{\vec{a}}} \left[\ln \left(f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a}) \right) \right] \Big|_{\vec{a} = \hat{\vec{a}}_{MAP}(\vec{r})} = \vec{0}$$

The estimation becomes the solving of those simultaneous equations.

(2) Measures of Errors

For the nonrandom variables $\vec{a}$, the first measure of interest is bias (vector) $\vec{b}(\vec{a})$:

$$\vec{b}(\vec{a}) \triangleq E\{\vec{a}_\varepsilon(\vec{r})\}$$
$$= E\{\hat{\vec{a}}(\vec{r})\} - \vec{a}$$

For multiple parameter estimation, if the bias vector $\vec{b}(\vec{a}) = \vec{0}$ (zero vector) we say that the corresponding $\vec{a}(\vec{r})$ is unbiased. Usually in the single parameter estimation, we may assume that the error $a_\varepsilon(\vec{r})$ is Gaussian, such that

$$f_{A_\varepsilon}(a_\varepsilon) = \frac{1}{\sqrt{2\pi}\,\sigma_{a_\varepsilon}} \exp\left[-\frac{a_\varepsilon^2}{2\sigma_{a_\varepsilon}^2}\right].$$

Analogous to the above argument, we can form the joint PDF for a set of K jointly Gaussian variables $\vec{a}_\varepsilon$ when the K-dimensional $\vec{a}$ is to be estimated such that

$$f_{\vec{A}_\varepsilon}(\vec{a}_\varepsilon) = \frac{1}{\sqrt{(2\pi)^K \det(\tilde{\Lambda}_\varepsilon)}} \exp\left[-\frac{1}{2}\vec{a}_\varepsilon^{\,T}\Lambda_\varepsilon^{-1}\vec{a}_\varepsilon\right]$$

where $\tilde{\Lambda}_\varepsilon \triangleq E\left\{(\vec{a}_\varepsilon - \bar{\vec{a}}_\varepsilon)(\vec{a}_\varepsilon^{\,T} - \bar{\vec{a}}_\varepsilon^{\,T})\right\}$

and $\bar{\vec{a}}_\varepsilon \triangleq E\{\vec{a}_\varepsilon\} = \vec{b}(\vec{a}) = E\{\hat{\vec{a}}(\vec{r})\} - \vec{a}$

Observing Eq. (247), we can draw equal-heigh[t] contours characterized by

$$\vec{a}_\varepsilon^{\,T} \tilde{\Lambda}_\varepsilon^{-1} \vec{a}_\varepsilon = C^2,$$

which is the ellipse equation for $k=2$.

$\dot{X}$ Concentration Ellipses

For $k=2$, the probability that the error vector $\vec{a}_\varepsilon$ lies inside an epllipse, whose equation is $\vec{a}_\varepsilon^{\,T} \tilde{\Lambda}_\varepsilon^{-1} \vec{a}_\varepsilon = C^2$, is

$$P = 1 - \exp\left(-\frac{C^2}{2}\right).$$

Proof: The area confined by $\vec{a}_\varepsilon^{\,T} \tilde{\Lambda}_\varepsilon^{-1} \vec{a}_\varepsilon = C^2$ is

$$A = \sqrt{\det(\tilde{\Lambda}_\varepsilon)}\ \pi C^2.$$

Thus, the differential area (the strip around the ellipse enclosing A) between the ellipses corresponding to C and $C+dC$ respetively is

$$dA = \sqrt{\det(\tilde{\Lambda}_\varepsilon)}\ 2\pi\ C\,dC.$$

The probability of $\vec{a}_\varepsilon$ in this differential area is

$$\frac{1}{2\pi\sqrt{\det(\tilde{\Lambda}_\varepsilon)}} \exp\left(-\frac{C^2}{2}\right).$$

Consequently, the probability for $\vec{a}_\varepsilon$ to lie out of this area is

$$1-P = \int_C^\infty x \exp\left(-\frac{x^2}{2}\right) dx$$

$$= \exp\left(-\frac{C^2}{2}\right).$$

Hence, $P = 1 - \exp\left(-\frac{C^2}{2}\right)$. #

Consequently, $\vec{a}_\varepsilon^{\,T} \tilde{\Lambda}_\varepsilon^{-1} \vec{a}_\varepsilon = C^2$ describes the concentration ellipses because it provides a measure of the concentration of the error density. A similar result holds for an arbitrary dimension K.

$$dv = \sqrt{\det(\tilde{\Lambda}_\varepsilon)} \; \frac{\sqrt{\pi^K}}{\Gamma\left(\frac{K}{2}+1\right)} \, K \, C^{K-1} \, dC$$

The value of the PDF on the ellipsoid is

$$\frac{1}{\sqrt{(2\pi)^K}\sqrt{\det(\tilde{\Lambda}_\varepsilon)}} \exp\left(-\frac{C^2}{2}\right).$$

Therefore

$$1 - P = \frac{K}{\sqrt{2^K}\,\Gamma\left(\frac{K}{2}+1\right)} \int_C^\infty x^{K-1} e^{-\frac{x^2}{2}} \, dx .$$

(3) Bounds on Estimation Errors for Nonrandom Variables

Property 1. Consider any unbiased estimate of a_i. Then

$$\sigma_{\varepsilon_i}^2 \triangleq \operatorname{var}\left[\hat{a}_i(\vec{R}) - a_i\right] \geq J^{ii}$$

where J^{ii} is the ii^{th} element in the $K \times K$ square matrix $\tilde{J}^{-1}$. The elements in $\tilde{J}$ are

$$J_{ij} \triangleq E\left\{ \frac{\partial \ln\left[f_{\vec{R}|\vec{A}}(\vec{R}\,|\,\vec{A})\right]}{\partial a_i} \times \frac{\partial \ln\left[f_{\vec{R}|\vec{A}}(\vec{R}\,|\,\vec{A})\right]}{\partial a_j} \right\}$$

$$= - E \left\{ \frac{\partial^2 \ln [f_{\vec{R}|\vec{A}} (\vec{r}|\vec{a})]}{\partial a_i \, \partial a_j} \right\}$$

or

$$\tilde{J} \triangleq E \left\{ \vec{\nabla}_{\vec{a}} \left(\ln [f_{\vec{R}|\vec{A}} (\vec{r}|\vec{a})] \right) \vec{\nabla}_{\vec{a}} \left(\ln [f_{\vec{R}|\vec{A}} (\vec{r}|\vec{a})] \right)^T \right\}$$

$$= - E \left\{ \vec{\nabla}_{\vec{a}} \left(\left[\vec{\nabla}_{\vec{a}} \left(\ln [f_{\vec{R}|\vec{A}} (\vec{r}|\vec{a})] \right) \right]^T \right) \right\}$$

$\tilde{J}$ matrix is commonly called the Fisher's Information Matrix. The equality in (258) holds if and only if

$$\hat{a}_i (\vec{r}) - a_i = \sum_{j=1}^{K} k_{ij} (\vec{a}) \frac{\partial \ln [f_{\vec{R}|\vec{A}} (\vec{r}|\vec{a})]}{\partial a_j}$$

$\forall \, a_i$ and $\forall \, \vec{r}$.

Proof:

Because $\hat{a}_i (\vec{r})$ is unbiased,

$$\int_{-\infty}^{\infty} [\hat{a}_i (\vec{r}) - a_i] \, f_{\vec{R}|\vec{A}} (\vec{r}|\vec{a}) \, d\vec{r} = 0$$

or

$$\int_{-\infty}^{\infty} \hat{a}_i (\vec{r}) \, f_{\vec{R}|\vec{A}} (\vec{r}|\vec{a}) \, d\vec{r} = a_i$$

Differentiating both sides with respect to a, we have

$$\int_{-\infty}^{\infty} \hat{a}_i(\vec{r}) \, \frac{\partial f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a})}{\partial a_j} \, d\vec{r}$$

$$= \int_{-\infty}^{\infty} \hat{a}_i(\vec{r}) \, \frac{\partial \ln[f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a})]}{\partial a_j} \, f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a}) \, d\vec{r}$$

$$= \delta_{ij} \quad \text{(Eq. 263)}$$

We define a (k+1)-dimensional vector

$$\vec{x} = \begin{bmatrix} \hat{a}_1(\vec{r}) - a_1 \\ \dfrac{\partial \ln[f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a})]}{\partial a_1} \\ \vdots \\ \dfrac{\partial \ln[f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a})]}{\partial a_k} \end{bmatrix}$$

The corresponding covariance matrix is

$$E\{\vec{x}\vec{x}^T\} = \left[\begin{array}{c|cccc} \sigma_{\epsilon_1}^2 & 1 & 0 & \cdots & 0 \\ \hline 1 & J_{11} & J_{12} & \cdots & J_{1K} \\ 0 & \vdots & \vdots & \ddots & \vdots \\ \vdots & & & & \\ 0 & J_{K1} & J_{K2} & \cdots & J_{KK} \end{array}\right]$$

$$\underbrace{\hphantom{J_{K1} \; J_{K2} \; \cdots \; J_{KK}}}_{\tilde{J}}$$

Since $E\{\vec{x}\vec{x}^T\} \geq 0$, $\det(E\{\vec{x}\vec{x}^T\}) \geq 0$.

$$\det(E\{\vec{x}\vec{x}^T\})$$

$$= \sigma_{\varepsilon_1}^2 \det(\tilde{J}) - \det\left(\begin{bmatrix} 1 & J_{12} & \cdots & J_{1k} \\ 0 & \vdots & & \\ \vdots & \vdots & & \\ 0 & J_{k2} & \cdots & J_{kk} \end{bmatrix}\right)$$

$$= \sigma_{\varepsilon_1}^2 \det(\tilde{J}) - \underbrace{\det\left(\begin{bmatrix} J_{22} & \cdots & J_{2k} \\ \vdots & & \\ J_{k2} & \cdots & J_{kk} \end{bmatrix}\right)}_{\text{cofactor } J_{11}}$$

(Cofactor J_{ij} is the determinant of the $(k-1)\times(k-1)$ submatrix where the i^{th} row and the j^{th} column have been removed from the $k\times k$ matrix $\tilde{J}$)

$$\therefore \quad \sigma_{\varepsilon_1}^2 \det(\tilde{J}) - \text{cofactor } J_{11} \geq 0.$$

If $\tilde{J}$ is nonsingular, we have

$$\sigma_{\varepsilon_1}^2 \geq \frac{\text{cofactor } J_{11}}{\det(\tilde{J})} = J^{11}.$$

Similarly, we can prove

$$\sigma_{\varepsilon_i}^2 \geq J^{ii}, \quad \forall i = 2, 3, \ldots, k. \quad \#$$

Property 2. Consider any unbiased estimate a. The concentration ellipse $\vec{a}_\varepsilon^{\,T} \tilde{\Lambda}_\varepsilon^{-1} \vec{a}_\varepsilon = C^2$ lies either outside or on the bound ellipse defined by $\vec{a}_\varepsilon^{\,T} \tilde{J} \vec{a}_\varepsilon = C^2$.

Property 3. The matrix

$$\tilde{\Lambda}_\varepsilon - \left\{ \vec{\nabla}_{\vec{a}} \left[\vec{g}_d^{\,T}(\vec{a}) \right] \right\}^T \tilde{J}^{-1} \left\{ \vec{\nabla}_{\vec{a}} \left[\vec{g}_d^{\,T}(\vec{a}) \right] \right\}$$

is nonnegative definite,

where $\vec{g}_d(\vec{a})$ specifies the estimate functions of the K parameters rather than the parameters themselves, such that

$$d_1 = g_{d_1}(\vec{a})$$
$$d_2 = g_{d_2}(\vec{a})$$
$$\vdots$$
$$d_M = g_{d_M}(\vec{a})$$

or $$\vec{d} = \vec{g}_d(\vec{a}) = \begin{bmatrix} g_{d_1}(\vec{a}) \\ g_{d_2}(\vec{a}) \\ \vdots \\ g_{d_M}(\vec{a}) \end{bmatrix}$$

$$d_{\varepsilon_i} = \hat{d}_i - g_{d_i}(\vec{a}), \quad i = 1, 2, \cdots, M.$$

and $$\tilde{\Lambda}_\varepsilon = E\{(\vec{d} - \vec{d}_\varepsilon)(\vec{d} - \vec{d}_\varepsilon)^T\}$$

$$\vec{d}_\varepsilon \triangleq \begin{bmatrix} d_{\varepsilon_1} \\ d_{\varepsilon_2} \\ \vdots \\ d_{\varepsilon_M} \end{bmatrix}, \quad \vec{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{bmatrix}$$

Property 4.

$$\mathrm{Var}(d_{\varepsilon_\ell}) \geq \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{\partial g_{d_\ell}(\vec{a})}{\partial a_i} J^{ij} \frac{\partial g_{d_\ell}(\vec{a})}{\partial a_j}.$$

Property 5. Assume that $\vec{g}_d(\vec{a}) \triangleq \tilde{G}_d \vec{a}$, where $\tilde{G}_d$ is an $M \times K$ matrix. If the estimates $\hat{\vec{d}}$ are unbiased, then

$$\tilde{\Lambda}_\varepsilon - \tilde{G}_d \tilde{J}^{-1} \tilde{G}_d^T \text{ is nonnegative}$$

definite

Property 6. Efficiency commutes with linear transformati[on] but not commutes with non linear transformat[ion]. In other words, if $\hat{\vec{a}}$ is efficient then $\hat{\vec{d}}$ will be efficient if and only if

$\vec{d} = \vec{g}_d(\vec{a})$ is a linear transformation.

X. Bounds on Estimation Errors: Multiple Random Parameters.

For random parameters, we have to defined a new information matrix $\tilde{J}_T$, which consists of two parts:

$$\tilde{J}_T \triangleq \tilde{J}_D + \tilde{J}_P ,$$

where $\tilde{J}_D$ is the Fisher's information matrix defined by Eq. (260) such that

$$\tilde{J}_D \triangleq E\left\{\left[\vec{\nabla}_{\vec{a}}\left(\ln\left[f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a})\right]\right)\right] \times \left[\vec{\nabla}_{\vec{a}}\left(\ln\left[f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a})\right]\right)\right]^T\right\}$$

$$= -E\left\{\vec{\nabla}_{\vec{a}}\left(\left[\vec{\nabla}_{\vec{a}}\left(\ln\left[f_{\vec{R}|\vec{A}}(\vec{r}|\vec{a})\right]\right)\right]^T\right)\right\}$$

represents the information obtained from the data $(\vec{r})$.

On the other hand, the matrix $\tilde{J}_P$ represents the a priori information such that

$$J_{P_{ij}} \triangleq E\left\{ \frac{\partial \ln[f_{\vec{A}}(\vec{a})]}{\partial a_i} \frac{\partial \ln[f_{\vec{A}}(\vec{a})]}{\partial a_j} \right\}$$

The mean-square errors are the diagonal elements in $\tilde{J}_T$. Three properties follow:

Property 1

$$\underbrace{E\{a_{\varepsilon_i}^2\}}_{\text{mean-square error}} \geq J_T^{ii}.$$

In other words, the diagonal elements in the inverse of the total information matrix $\tilde{J}_T$ are the lower bounds on the corresponding mean-square errors. The correlation matrix of the errors is

$$\tilde{R}_\varepsilon \triangleq E\{\vec{a}_\varepsilon \vec{a}_\varepsilon^T\},$$

where the diagonal elements in $\tilde{R}_\varepsilon$ represent the mean-square errors $E\{a_{\varepsilon_i}^2\}$.

Property 2. The matrix $\tilde{J}_T - \tilde{R}_\varepsilon^{-1}$ is nonnegative definite.

Property 3. If $\tilde{J}_T = \tilde{R}_\varepsilon^{-1}$, all of the estimates are efficient (achieve minimum variances). A necessary and sufficient condition for this to be true is that $f_{\vec{A}|\vec{R}}(\vec{a}|\vec{r})$ be Gaussian $\forall \vec{r}$. This will be true if $\tilde{J}_T$ is constant. (Quizz 3)

A special case of interest occurs when the a priori density is a k^{th}-order Gaussian density. Then

$$\tilde{J}_P = \tilde{\Lambda}_a^{-1},$$

where $\tilde{\Lambda}_a$ is the covariance matrix of the random parameters $\vec{a}$ such that

$$\tilde{\Lambda}_a \triangleq E\left\{(\vec{a} - E\{\vec{a}\})(\vec{a} - E\{\vec{a}\})^T\right\}.$$

If $\vec{a}$ constitutes an independent Gaussian process, then $\tilde{\Lambda}_a$ is diagonal and

$$J_{P_{ij}} = \frac{1}{\sigma_{a_i}^2} \delta_{ij},$$

where $\tilde{J}_P$ is diagonal and $\sigma_{a_i}^2 \triangleq E\{(a_i - E(a_i))^2\}$

Results similar to Properties 3–6 for nonrandom parameters can also be derived for the random parameter case.

2.4.4 Summary of Estimation Theory

Read by Yourselves!