EE 631: Estimation and Detection
Part 4

Dr. Mehrdad Soumekh

Parameter Estimation

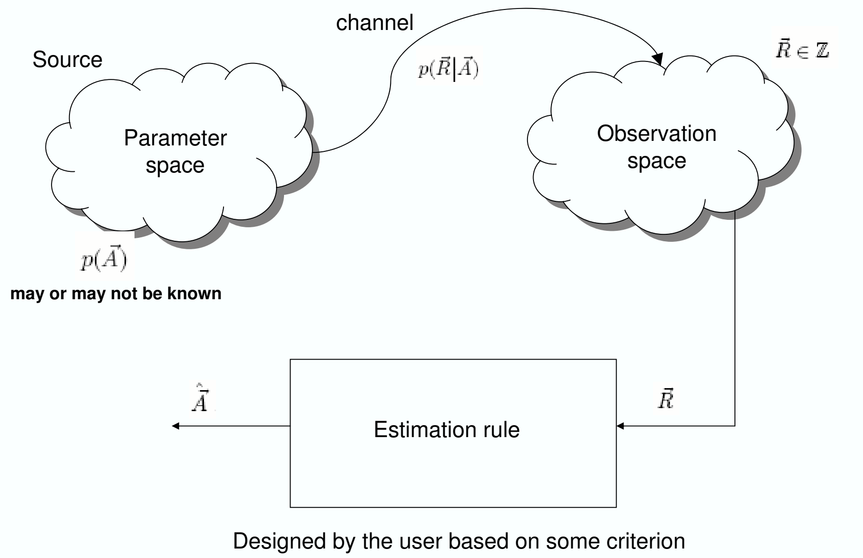

Figure 1. Model assumed in parameter estimation problem

- We are interested in the value of the parameter $\vec{A}$
- Assign a cost function to the estimator error as a measure of our receiver performance.
- Transition pdf of observed/measured vector is given by $p(\vec{R}|\vec{A})$

Estimation error : $\vec{\epsilon}(\vec{R}) = \vec{A} - \hat{\vec{A}}(\vec{R})$
Therefore : $\epsilon_1 = \hat{\vec{A}} - \hat{A}_1$, $\epsilon_2 = \hat{\vec{A}} - \hat{A}_2$ etc.
i.e. ϵ depends on the observation $\vec{R}$ and the estimation rule.

Cost function: $\mathfrak{R}(\vec{\epsilon})$
We try to minimize $\mathfrak{R}$ over the probability space of $\vec{A}$ and $\vec{R}$ by selecting the appropriate estimation rule.

Single Parameter Case
Assume apriori pdf $P(A)$ is known.

1. Minimum Mean Squared Error (MMSE) estimator:

$$\begin{aligned}\mathfrak{R}(\vec{\epsilon}) &= E\left[|\vec{\epsilon}|^2\right] \\ &= E\left[(A - \hat{A})^2\right] \\ &= E\left[(A - \hat{A}(R))^2\right]\end{aligned}$$

where the expectation is taken over the probability space of $\vec{A}$ and $\vec{R}$.

Objective: Choose $\hat{A}(\vec{R})$ that minimizes the above cost function.

$$\begin{aligned}\Re(\epsilon) &= \int_{\mathbb{Z}}\int_A \left[(A-\hat{A}(R))^2 p(A,\vec{R})dA.d\vec{R}\right] \\ &= \int_{\mathbb{Z}} \underbrace{\int_A \left[(A-\hat{A}(R))^2 p(A|\vec{R})dA\right]}_{\Re_1(\epsilon)} \underbrace{p(\vec{R})}_{+ve} d\vec{R}\end{aligned}$$

Thus, to minimize $\Re$, it is sufficient to minimize the inner integral:

$$\Re_1(\epsilon) = \int_A (A-\hat{A}(\vec{R}))^2 p(A|\vec{R})dA$$

For this, we find the minimum by obtaining the derivative of $\Re_1$ with respect to $\hat{A}$ and setting it equal to zero.

$$\begin{aligned}\frac{\delta\Re_1}{\delta\hat{A}}(\vec{\epsilon}) &= 2\big(\hat{A}(\vec{R})\big)p(A|\vec{R})dA\Big|_{\hat{A}=\hat{A}_{MMSE}} \\ \text{or } \int_A \hat{A}_{MMSE}(\vec{R})p(A|\vec{R})dA &= \int_A A.p(A|\vec{R})dA \\ \Rightarrow \hat{A}_{MMSE}(\vec{R})\underbrace{\int_A p(A|\vec{R})dA}_{=1} &= E(A|\vec{R}) \\ \Rightarrow \hat{A}_{MMSE}(\vec{R}) &= E(A|\vec{R})\end{aligned}$$

2. Minimum absolute value of error cost function:

$$\begin{aligned}\Re(\vec{\epsilon}) &\triangleq E\big[|A-\hat{A}(\vec{R})|\big] \\ &= \int_{\mathbb{Z}}\int_A |A-\hat{A}(\vec{R})|p(A,\vec{R})dA.d\vec{R} \\ &= \int_{\mathbb{Z}}\underbrace{\int_A |A-\hat{A}(\vec{R})|p(A|\vec{R})dA}_{\Re_1(\epsilon)}\, p(\vec{R})d\vec{R}\end{aligned}$$

Thus the minimization can be performed on:

$$\begin{aligned}\Re(\vec{\epsilon}) &\triangleq \int_{-\infty}^{\infty} |A-\hat{A}(\vec{R})|p(A|\vec{R})dA \\ &= \int_{-\infty}^{\hat{A}(\vec{R})} \big[\hat{A}(\vec{R})-A\big]p(A|\vec{R})dA + \int_{\hat{A}(\vec{R})}^{\infty} \big[A-\hat{A}(\vec{R})\big]p(A|\vec{R})dA\end{aligned}$$

To minimize, use:

$$\frac{\delta\Re_1}{\delta\hat{A}}(\vec{\epsilon})\Big|_{\hat{A}=\hat{A}_{abs}} = 0$$

$$\Rightarrow \int_{-\infty}^{\hat{A}} p(A|\vec{R})dA - \int_{\hat{A}}^{\infty} p(A|\vec{R})dA + \hat{A}p(\hat{A}|\vec{R}) - \hat{A}p(\hat{A}|\vec{R})\Big|_{\hat{A}=\hat{A}_{abs}} = 0$$

$$\Rightarrow \int_{-\infty}^{\hat{A}_{abs}} p(A|\vec{R})dA = \int_{\hat{A}_{abs}}^{\infty} p(A|\vec{R})dA$$

Thus $\hat{A}_{abs}$ = median of $(A|\vec{R})$.

3. Maximum A posteriori (MAP) estimator:

$\hat{A}_{MAP}$ is the point where $P(A|\vec{R})$, i.e the a posteriori pdf achieves its maximum.
Since ln[.] is a monotone increasing transformation, the max point of $\ln[p(a|\vec{R})]$ is also the $\hat{A}_{MAP}(\vec{R})$, i.e.

$$\frac{\delta}{\delta A}\ln[p(A|\vec{R})]\Big|_{A=\hat{A}_{MAP}} = 0$$

From the Bayes equation, we have:

$$p(A|\vec{R}) = \frac{p(\vec{R}|A)p(A)}{p(\vec{R})}$$

Use this in the MAP equation:

$$\frac{\delta}{\delta A}\ln p(\vec{R}|A) + \frac{\delta}{\delta A}\ln p(A) - \underbrace{\frac{\delta}{\delta A}\ln p(\vec{R})}_{=0 \text{ since } \vec{R} \text{ is invariant in } A}\Big|_{\hat{A}=\hat{A}_{MAP}} = 0$$

The MAP equation becomes:

$$\frac{\delta}{\delta A}\ln p(\vec{R}|A) + \frac{\delta}{\delta A}\ln p(A)\Big|_{\hat{A}=\hat{A}_{MAP}} = 0$$

If the function $P(A)$ is a relatively smooth function as compared with $p(\vec{R}|A)$, then the MAP estimator can be approximated by:

$$\frac{\delta}{\delta A}\ln p(\vec{R}|A)\Big|_{\hat{A}=\hat{A}_{MAP}} = 0$$

This is also true the other way round. Therefore, if either $p(\vec{R}|A)$ or $p(A)$ is a flat function with respect to A, then the MAP estimate point is dictated by the other pdf.

Maximum likelihood (ML) estimate:

ML estimate is used when the parameters to be estimated are non random, i.e. a priori pdf $p(A)$ is not available. Define the ML estimate as the one that maximizes the likelihood function

$$\mathfrak{L}(\vec{R}, A) = p(\vec{R}, A)$$

$p(\vec{R}, A)$ is used because the parameters are non-random. The equivalent for $p(\vec{R}, A)$ in the case random A is $p(\vec{R}|A)$.
Note that for the random case we have:

$$p(\vec{R}, A) = p(\vec{R}|A)p(A)$$

If not known, then $p(A)$ is assumed to be a "flat" or constant valued distribution. Therefore, the ML estimate for the non-random parameter case is the same as the MAP estimate for the random parameter case when $p(A)$ is flat.

Properties of an estimator

- An estimate of A is unbiased if

$$E[\hat{A}(\vec{R})] = A$$

- A biased estimate has the form:

$$E[\hat{A}(\vec{R})] = A + \bar{b}_{\hat{A}}$$

where $\bar{b}_{\hat{A}} = E[\hat{A}(\vec{R}) - A]$ is called the *bias*. The bias depends on:
i) Measurement
ii) Class of estimator that is used

Desirable properties of an estimator

- The estimator should be unbiased.
- The variance for the estimate should be the minimum possible. An optimum estimate is the unbiased minimum variance estimate (UMVE). - Variance of the estimate goes to zero when the number of observations goes to infinity. i.e.

$$\lim_{N\to\infty} Var[\hat{A}(\vec{R})] = 0$$

In this case, $\hat{A}$ is called a consistent estimate.

Fisher information

Likelihood function:

$$\mathfrak{L}(a) \triangleq p(\vec{R}; a)$$

$$\Rightarrow \ln \mathfrak{L}(a) = \ln p(\vec{R}; a)$$

Score is identified by:

$$\mathbb{V} = \frac{\delta}{\delta a} \ln \mathfrak{L}(a) = \frac{\delta}{\delta a} \ln p(\vec{R}; a)$$

$$= \frac{p'(\vec{R};a)}{p(\vec{R};a)} = \frac{\frac{\delta}{\delta a} p(\vec{R};a)}{p(\vec{R};a)}$$

Note: The knowledge of $\mathbb{V}$ is equivalent to knowing the likelihood function. Thus the score $\mathbb{V}$ is sufficient statistic in this case.

Properties of Score

i) Mean:

$$\begin{aligned} E(\mathbb{V}) &= \int_{\mathbb{Z}} \frac{\frac{\delta}{\delta a} p(\vec{R};a)}{p(\vec{R};a)} p(\vec{R}; a) d\vec{R} \\ &= \int_{\mathbb{Z}} \frac{\delta}{\delta a} p(\vec{R}; a) d\vec{R} \\ &= \frac{\delta}{\delta a} \int_{\mathbb{Z}} \underbrace{p(\vec{R}; a) d\vec{R}}_{=1} \\ &= \frac{\delta}{\delta a}(1) = 0 \\ \Rightarrow E(\mathbb{V}) &= 0 \end{aligned}$$

ii) Variance of score:

$$\begin{aligned} I_{\vec{R}}(a) &= var(\mathbb{V}) = E(\mathbb{V}^2) \\ &= \int_{\mathbb{Z}} \left[\frac{\delta}{\delta a} \ln p(\vec{R}; a)\right]^2 p(\vec{R}; a) d\vec{R} \end{aligned}$$

Note that the variance of score is only a function of the parameter a. The subscript $\vec{R}$ in $I_{\vec{R}}$ is used to identify the channel that is used for observation. In other words, this is for the class or set of measurements from a specific channel.
$I_{\vec{R}}(a)$ is called the Fisher information and it depends on the specific channel that is used for the observation. However, it is invariant in a specific $\vec{R}$. Let

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

where r_i's are independent. Thus, we have:

$$\begin{aligned} p(\vec{R};a) &= \prod_{i=1}^{N} p(r_i;a) \\ \Rightarrow \ln p(\vec{R};a) &= \sum_{i=1}^{N} \ln p(r_i;a) \\ \Rightarrow \frac{\delta}{\delta a} \ln p(\vec{R};a) &= \sum_{i=1}^{N} \frac{\delta}{\delta a} \ln p(r_i;a) \end{aligned}$$

If we denote the individual score for the ith measurement as:

$$\mathbb{V}_i \triangleq \frac{\delta}{\delta a} \ln p(r_i;a)$$

then

$$\begin{aligned} \mathbb{V} &= \frac{\delta}{\delta a} \ln p(\vec{R};a) = \sum_{i=1}^{N} \mathbb{V}_i \\ \Rightarrow \mathbb{V}(\vec{R}) &= \sum_{i=1}^{N} \mathbb{V}_i(r_i) \end{aligned}$$

Since r_i's are independent, $\mathbb{V}_i$'s are also independent rv's. Therefore, we have:

$$Var(\mathbb{V}) = \sum_{i=1}^{N} Var(\mathbb{V}_i)$$

Also, variance of score is the Fisher information; thus

$$I_{\vec{R}}(a) = \sum_{i=1}^{N} I_{r_i}(a)$$

This implies that Fisher information is additive for independent observations.

If r_i's are independent identically distributed (iid) with pdf $p_0(r_i)$, then

$$\begin{aligned} \mathbb{V}_i &= \frac{\delta}{\delta a} \ln p_0(r_i) \\ \Rightarrow I_i(a) &= Var(\mathbb{V}_i) \triangleq I_0(a) \\ \Rightarrow I_{\vec{R}}(a) &= \sum_{i=1}^{N} I_i(a) = N.I_0(a) \end{aligned}$$

Another expression for Fisher information

We begin with

$$\begin{aligned} \frac{\delta}{\delta a}\mathbb{V} &= \frac{\delta^2}{\delta a^2} \ln p(\vec{R};a) = \frac{\delta}{\delta a}\left(\frac{p'}{p}\right) \\ &= \frac{p''p-(p')^2}{p^2} = \frac{p''}{p} - \left(\frac{p'}{p}\right)^2 \end{aligned}$$

Consider

$$\begin{aligned} E\left(\frac{p''}{p}\right) &= \int_{\mathbb{Z}} \frac{\frac{\delta^2}{\delta a^2} p(\vec{R};a)}{p(\vec{R};a)} p(\vec{R};a) d\vec{R} \\ &= \int_{\mathbb{Z}} \frac{\delta^2}{\delta a^2} p(\vec{R};a) d\vec{R} \\ &= \frac{\delta^2}{\delta a^2} \underbrace{\int_{\mathbb{Z}} p(\vec{R};a) d\vec{R}}_{=1} \\ &= 0 \end{aligned}$$

Since $\frac{\delta^2}{\delta a^2} p(\vec{R};a) = \frac{p''}{p} - \left(\frac{p'}{p}\right)^2$, the expectation on both side yields:

$$\begin{aligned} E\Big[\tfrac{\delta^2}{\delta a^2} p(\vec{R};a)\Big] &= E\Big[-\left(\tfrac{p'}{p}\right)^2\Big] \\ &= -E[\mathbb{V}^2] \\ &= I_{\vec{R}}(a) \\ \Rightarrow I_{\vec{R}}(a) &= -E\Big[\tfrac{\delta^2}{\delta a^2} p(\vec{R};a)\Big] \\ &= -E\Big[\tfrac{\delta}{\delta a}\mathbb{V}\Big] \end{aligned}$$

Cramer Rao Bound

An estimate $\hat{a_1}(\vec{R})$ is said to be more *"efficient"* than another estimate $\hat{a_1}(\vec{R})$ if

$$E\Big[[a - \hat{a_1}(\vec{R})]^2\Big] < E\Big[[a - \hat{a_2}(\vec{R})]^2\Big]$$

Let $\hat{a}(\vec{R})$ be an estimate of a, e.g.

$$\hat{a}(\vec{R}) = a + \underbrace{b_{\hat{a}}(\vec{R})}_{bias}$$

We know that the mean value of the score is zero, i.e. $E(\mathbb{V}) = 0$.
Thus, for the covariance of the estimator and score, we have:

$$E\Big[(\mathbb{V} - \bar{\mathbb{V}})(\hat{a} - \bar{\hat{a}})\Big] = E(\mathbb{V}\hat{a}) - \bar{\hat{a}}E(\mathbb{V}) = E(\mathbb{V}\hat{a})$$

Substitute for $\mathbb{V}$ on the right side:

$$\begin{aligned} cov(\mathbb{V},\hat{a}) &= E(\mathbb{V}.\hat{a}) \\ &= E\Big[\tfrac{\delta}{\delta a}\ln p(\vec{R};a).\hat{a}(\vec{R})\Big] \\ &= \textstyle\int_{\mathbb{Z}} \frac{\frac{\delta}{\delta a} p(\vec{R};a)}{p(\vec{R};a)}.\hat{a}(\vec{R}) p(\vec{R};a) d\vec{R} \\ &= \textstyle\int_{\mathbb{Z}} \frac{\delta}{\delta a} p(\vec{R};a).\hat{a}(\vec{R}) d\vec{R} \end{aligned}$$

Since $\vec{R}$ and as a result $\hat{a}(\vec{R})$ are invariant in a, we can move the $\frac{\delta}{\delta a}$ outside:

$$\begin{aligned} cov(\mathbb{V},\hat{a}) &= \tfrac{\delta}{\delta a} \underbrace{\int_{\mathbb{Z}} \hat{a}(\vec{R}) p(\vec{R};a) d\vec{R}}_{\text{expected value of the estimate}} \\ &= \tfrac{\delta}{\delta a}[a + b_{\hat{a}}] \\ &= 1 + f_{\hat{a}}(a) \end{aligned}$$

where $f_{\hat{a}}(a) = \frac{\delta}{\delta a} b_{\hat{a}}$. Note that $f_{\hat{a}}(a)$ is not an rv. Also, for an unbiased estimator $f_{\hat{a}}(a) = 0$ and $cov = 1$.

Moreover, from Schwartz inequality, we have:

$$var(\hat{a}).var(\mathbb{V}) \geq cov^2(\mathbb{V},\hat{a})$$

Note that

$$E[(x+\alpha y)^2] = E(x^2) + E(y^2) + \alpha^2 + 2E(xy)\alpha \geq 0$$

Consider the quadratic in α

$$E(y^2).\alpha^2 + \alpha.2E(xy) + E(x^2) = 0$$

For this to have no solution,

$$E^2(xy) \quad < E(x^2)E(y^2)$$

$$\Rightarrow var(\hat{a}) \geq \frac{cov^2(\mathbb{V},\hat{a})}{var(\mathbb{V})}$$

$$\Rightarrow var(\hat{a}) \geq \frac{[1+f_{\hat{a}}(a)]^2}{I_{\vec{R}}(a)}$$

This is called the **Cramer Rao bound** which shows how Fisher information limits the performance of estimators. e.g. For an unbiased $E(\hat{a}) = \bar{\hat{a}} = a$

$$\begin{aligned} var(\hat{a}) \quad &= E\Big[(\hat{a}-\bar{\hat{a}})^2\Big] \\ &= E\Big[(\hat{a}-a)^2\Big] \triangleq \text{mean square error} \end{aligned}$$

Also for an unbiased estimator, $f_{\hat{a}}(a) = 0$. Therefore, the CR bound for unbiased information is given by:

$$MSE(\hat{a}) \geq \frac{1}{I_{\vec{R}}(a)}$$

The efficiency of an unbiased estimator is defined by:

$$eff(\hat{a}) = \frac{\frac{1}{I_{\vec{R}}(a)}}{var(\hat{a})} \leq 1$$

It is to be noted that an ML estimator is not necessarily an *efficient* one. However, if one such estimator exists, then it must be ML.